ANALYZING EXPERIMENTAL DATA BY REGRESSION x SAS:
Annotated Outputs

BU-705-M May 1980

S. Aref and W. W. Piegorsch
Biometrics Unit, Cornell University, Ithaca, New York

Abstract

This annotated output was created to assist the student in learning how to use the SAS computer package. It is intended as an aid to the student in a course covering the methods of data analysis, as presented in Analyzing Experimental Data by Regression by D.M. Allen and F.B. Cady.

SAS programs for nine data sets from Allen and Cady's book were developed. They were tied as closely to the text as possible and were written to show several different methods of analysis. Hence, the programs do not always demonstrate the most efficient and logical way of doing the data analysis. Rather, they illustrate different aspects of the methods presented in the book.

BU-705-M in the Biometrics Unit Mimeo Series, Cornell University, Ithaca, New York.

TABLE OF CONTENTS

*D.M. Allen and F.B. Cady, Analyzing Experimental Data by Regression.

For SAS programs the following version was used:
SAS User's Guide (1979). J.T. Helwig and K.A. Council, editors.
SAS Institute, Inc., Raleigh, North Carolina.

NOTE: THE JOB VMPJQ001 HAS BEEN RUN UNDER RELEASE 79.2B OF SAS AT CORNELL UNIVERSITY.

```
2          DATA ARSENIC;  (UNITS 4,5,6)
3               INPUT X1 Y;
4               X0=1; Set a column of 1's only for printing the X matrix. SAS-GLM supplies
5               CARDS;                                                    its own.
```

NOTE: DATA SET WORK.ARSENIC HAS 10 OBSERVATIONS AND 3 VARIABLES. 465 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.21 SECONDS AND 116K.

```
16          TITLE ARSENIC DATA;
17    (A)   PROC PRINT;                Print X-matrix for mean model.
18          VARIABLES X0;
```

NOTE: THE PROCEDURE PRINT USED 0.13 SECONDS AND 116K AND PRINTED PAGE 1.

```
19          PROC GLM;                          P option prints yi, ŷi, (yi-ŷi) for each i.
20    (B)   MODEL Y=X0/NOINT P;  Regress Y on X0 (no intercept for mean model; we supply our own,
21          OUTPUT OUT=NEW1                                                          i.e. X0)
22               RESIDUAL=RESID1     PREDICTED=YHAT1;
```

The OUTPUT statement must be used to create a data set with residual & predicted values.

NOTE: DATA SET WORK.NEW1 HAS 10 OBSERVATIONS AND 5 VARIABLES. 296 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.30 SECONDS AND 184K AND PRINTED PAGE 2.

```
23    (C)   PROC PLOT;
24          PLOT RESID1*YHAT1;  Plot residuals on Y-axis vs. predicted on X-axis.
```

NOTE: THE PROCEDURE PLOT USED 0.20 SECONDS AND 124K AND PRINTED PAGE 3.

```
25    (D)   PROC PRINT;
26          VARIABLES X0 X1;  Print X-matrix for intercept & slope model.
```

NOTE: THE PROCEDURE PRINT USED 0.16 SECONDS AND 116K AND PRINTED PAGE 4.

```
27          PROC GLM;
28    (E)   MODEL Y=X1/P;  Regress Y on X1 (now we let SAS-GLM supply intercept variable),
29          OUTPUT OUT=NEW2
30               RESIDUAL=RESID2     PREDICTED=YHAT2;  } OUTPUT statement necessary for
                                                          residual plot.
```

NOTE: DATA SET WORK.NEW2 HAS 10 OBSERVATIONS AND 7 VARIABLES. 217 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.31 SECONDS AND 186K AND PRINTED PAGE 5.

```
31    (F)   PROC PLOT;
32          PLOT RESID2*YHAT2;  Residual plot of intercept & slope model.
```

NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 124K AND PRINTED PAGE 6.

```
33    (G)   PROC GLM;                    Regress Y on X0 and X1 (no intercept, we supply X0)
34          MODEL Y=X0 X1/NOINT P;
```

NOTE: THE PROCEDURE GLM USED 0.27 SECONDS AND 162K AND PRINTED PAGE 7.

```
35    (H)   PROC GLM;                    Regress Y on X1 and X0 (again, we supply X0)
36          MODEL Y=X1 X0/NOINT P;
```

NOTE: THE PROCEDURE GLM USED 0.26 SECONDS AND 162K AND PRINTED PAGE 8.

Don't need residual plot since regression equation will be the same as Y on X1 with intercept. Therefore, we don't need an OUTPUT statement.

(A)

ARSENIC DATA

OBS	X0
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1

X matrix for mean model

(B)

ARSENIC DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y Regress Y on X_0 (no intercept)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL *Regression*	1	26.50384000	26.50384000	25.86	0.0007
ERROR *Residuals*	9	9.22296000	1.02477333 ← estimated s^2 (compare with NOTES: p. 4.5)		
UNCORRECTED TOTAL	10	35.72680000			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.741848	62.1813	1.01231089	1.62800000 = $\bar{y}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X0	1	26.50384000	25.86	0.0007

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X0	1	26.50384000	25.86	0.0007

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X0	1.62800000 = $\hat{\mu}$	5.09	0.0007	0.32012081

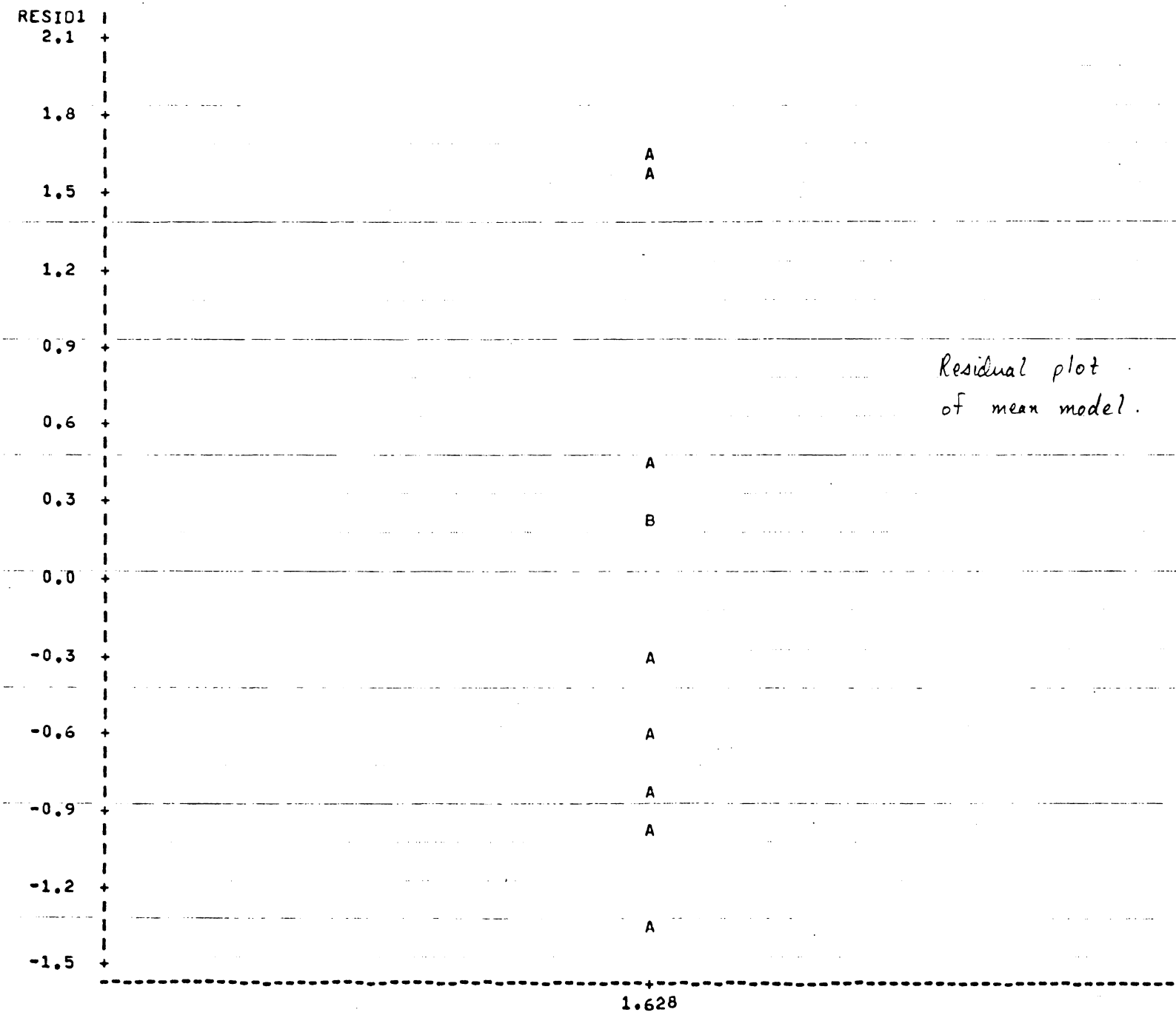
OBSERVATION	OBSERVED VALUE y_i	PREDICTED VALUE $\hat{y}_i = \bar{y}$	RESIDUAL $y_i - \hat{y}_i = y_i - \bar{y}$
1	3.19000000	1.62800000	1.56200000
2	3.26000000	1.62800000	1.63200000
3	1.82000000	1.62800000	0.19200000
4	1.02000000	1.62800000	-0.60800000
5	1.85000000	1.62800000	0.22200000
6	2.05000000	1.62800000	0.42200000
7	1.34000000	1.62800000	-0.28800000
8	0.79000000	1.62800000	-0.83800000
9	0.66000000	1.62800000	-0.96800000
10	0.30000000	1.62800000	-1.32800000

SUM OF RESIDUALS	0.00000000 ← Should sum to zero.
SUM OF SQUARED RESIDUALS	9.22296000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	0.53355929
DURBIN-WATSON D	0.47712448

(c)

ARSENIC DATA

PLOT OF RESID1*YHAT1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot of mean model.

(D)

ARSENIC DATA

OBS	X0	X1
1	1	2
2	1	4
3	1	8
4	1	10
5	1	12
6	1	15
7	1	21
8	1	23
9	1	30
10	1	36

X-matrix for intercept & slope model

(E)

ARSENIC DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X_1 with intercept.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	1	6.88257399	6.88257399	23.53	0.0013
ERROR	8	2.34038601	0.29254825		
CORRECTED TOTAL	9	9.22296000			

Note ↑

R-SQUARE	C.V.	STD DEV	Y MEAN
0.746244	33.2234	0.54087730	1.62800000 $= \bar{y}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	6.88257399	23.53	0.0013

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	1	6.88257399	23.53	0.0013

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	$b_0 = 2.88622593$	9.29	0.0001	0.31071996
X1	$b_1 = -0.07815068$	-4.85	0.0013	0.01611225

↖ Compare with NOTES: pp. 5.9-10.

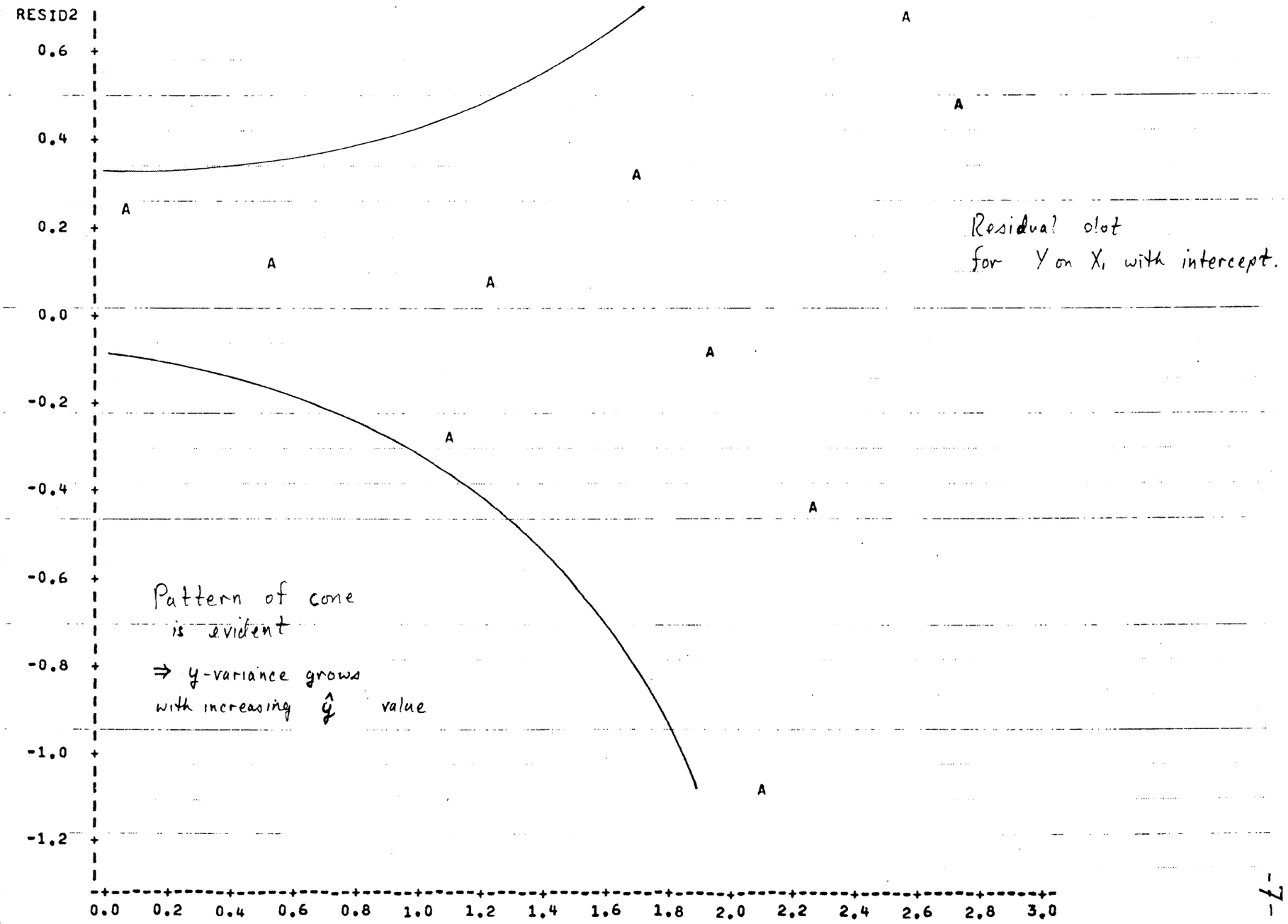
OBSERVATION	OBSERVED VALUE y_i	PREDICTED VALUE $\hat{y}_i$	RESIDUAL $y_i - \hat{y}_i$
1	3.19000000	2.72992457	0.46007543
2	3.26000000	2.57362321	0.68637679
3	1.82000000	2.26102050	-0.44102050
4	1.02000000	2.10471914	-1.08471914
5	1.85000000	1.94841778	-0.09841778
6	2.05000000	1.71396575	0.33603425
7	1.34000000	1.24506167	0.09493833
8	0.79000000	1.08876032	-0.29876032
9	0.66000000	0.54170556	0.11829444
10	0.30000000	0.07280149	0.22719851

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	2.34038601
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.23937056
DURBIN-WATSON D	1.40876096

(F)

ARSENIC DATA

PLOT OF RESID2*YHAT2 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



ARTIFICIAL DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X_0 and X_1 (no intercept)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	33.38641399	16.69320699	57.06	0.0001
ERROR	8	2.34038601	0.29254825		
UNCORRECTED TOTAL	10	35.72680000			

Compare with NOTES: p.6.5

R-SQUARE	C.V.	STD DEV	Y MEAN
0.934492	33.2234	0.54087730	1.62800000 = $\bar{Y}$

SOURCE	DF	TYPE (I) SS	F VALUE	PR > F
X0 alone	1	26.50384000	90.60	0.0001
X1 after X_0	1	6.88257399	23.53	0.0013

Compare with NOTES: p.6.5

SOURCE	DF	TYPE (IV) SS	F VALUE	PR > F
X0 after X_1	1	25.24177252	86.28	0.0001
X1 after X_0	1	6.88257399	23.53	0.0013

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X0	b_0 = 2.88622593	9.29	0.0001	0.31071996
X1	b_1 = -0.07815068	-4.85	0.0013	0.01611225

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	3.19000000	2.72992457	0.46007543
2	3.26000000	2.57362321	0.68637679
3	1.82000000	2.26102050	-0.44102050
4	1.02000000	2.10471914	-1.08471914
5	1.85000000	1.94841778	-0.09841778
6	2.05000000	1.71396575	0.33603425
7	1.34000000	1.24506167	0.09493833
8	0.79000000	1.08876032	-0.29876032
9	0.66000000	0.54170556	0.11829444
10	0.30000000	0.07280149	0.22719851

same as Y on X_1 with intercept.

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	2.34038601
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.23937056
DURBIN-WATSON D	1.40876096

(H)

ARSENIC DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X_1 and X_0 (no intercept)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	33.38641399	16.69320699	57.06	0.0001
ERROR	8	2.34038601	0.29254825		
UNCORRECTED TOTAL	10	35.72680000			

Compare with NOTES: p.6.6.

R-SQUARE	C.V.	STD DEV	Y MEAN
0.934492	33.2234	0.54087730	1.62800000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1 alone	1	8.14464146	27.84	0.0007
X0 after X_1	1	25.24177252	86.28	0.0001

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1 after X_0	1	6.88257399	23.53	0.0013
X0 after X_1	1	25.24177252	86.28	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	$b_1 = -0.07815068$	-4.85	0.0013	0.01611225
X0	$b_0 = 2.88622593$	9.29	0.0001	0.31071996

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	3.19000000	2.72992457	0.46007543
2	3.26000000	2.57362321	0.68637679
3	1.82000000	2.26102050	-0.44102050
4	1.02000000	2.10471914	-1.08471914
5	1.85000000	1.94841778	-0.09841778
6	2.05000000	1.71396575	0.33603425
7	1.34000000	1.24506167	0.09493833
8	0.79000000	1.08876032	-0.29876032
9	0.66000000	0.54170556	0.11829444
10	0.30000000	0.07280149	0.22719851

Same as preceeding two GLM proceedures.

SUM OF RESIDUALS	-0.00000000
SUM OF SQUARED RESIDUALS	2.34038601
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	0.23937056
DURBIN-WATSON D	1.40876096

```
37          DATA FIREFLY;          (UNITS 10,11,13)
38               INPUT Y X1 X2;
39               X0=1;
40               CARDS;

NOTE: DATA SET WORK.FIREFLY HAS 17 OBSERVATIONS AND 4 VARIABLES. 361 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.07 SECONDS AND 116K.

58          TITLE FIREFLY DATA;
59     (A)  PROC PRINT;
60               VAR X0 X1 X2;          Print X-matrix.

NOTE: THE PROCEDURE PRINT USED 0.14 SECONDS AND 116K AND PRINTED PAGE 9.

61     (B)  PROC GLM;
62               MODEL Y=X1/P;          Regress Y on X1 with intercept.
63               OUTPUT OUT=NEW3  RESIDUAL=RESID3   PREDICTED=YHAT3;  } OUTPUT statement for
                                                                        residual plot.
NOTE: DATA SET WORK.NEW3 HAS 17 OBSERVATIONS AND 6 VARIABLES. 250 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.35 SECONDS AND 184K AND PRINTED PAGES 10 TO 11.

64     (C)  PROC PLOT;
65               PLOT RESID3*YHAT3;     Residual plot for  ŷ = b0x0 + b1x1 .

NOTE: THE PROCEDURE PLOT USED 0.20 SECONDS AND 124K AND PRINTED PAGE 12.

66          PROC GLM;
67     (D)       MODEL Y=X1 X2/P;       Regress Y on X1 and X2 with intercept.
68               OUTPUT OUT=NEW4  RESIDUAL=RESID4   PREDICTED=YHAT4;

NOTE: DATA SET WORK.NEW4 HAS 17 OBSERVATIONS AND 8 VARIABLES. 191 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.32 SECONDS AND 186K AND PRINTED PAGES 13 TO 14.

69     (E)  PROC PLOT;
70               PLOT RESID4*YHAT4;     Residual plot of  ŷ = b0x0 + b1x1 + b2x2 .

NOTE: THE PROCEDURE PLOT USED 0.24 SECONDS AND 124K AND PRINTED PAGE 15.

71     (F)  PROC GLM;                   Regress Y on X2 and X1 with intercept.
72               MODEL Y=X2 X1/P;       (residual plot same as Y on X1 and X2 ⇒ don't need OUTPUT)

NOTE: THE PROCEDURE GLM USED 0.27 SECONDS AND 162K AND PRINTED PAGES 16 TO 17.

73     (G)  PROC PLOT;          Plot variables on each other.
74               PLOT X2*X1 Y*X1 Y*X2;      (NOTE: this is 3 separate plots).

NOTE: THE PROCEDURE PLOT USED 0.38 SECONDS AND 134K AND PRINTED PAGES 18 TO 20.

75          PROC GLM;                Regress Y on X2 with intercept.
76     (H)       MODEL Y=X2/P;
77               OUTPUT OUT=NEW5  RESIDUAL=RESID5   PREDICTED=YHAT5;

NOTE: DATA SET WORK.NEW5 HAS 17 OBSERVATIONS AND 10 VARIABLES. 155 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.38 SECONDS AND 184K AND PRINTED PAGES 21 TO 22.

78          PROC PLOT;
79     (J)       PLOT RESID5*YHAT5;     Residual plot of  ŷ = b0x0 + b2x2 .

NOTE: THE PROCEDURE PLOT USED 0.25 SECONDS AND 124K AND PRINTED PAGE 23.
```

(A)

FIREFLY DATA

OBS	X0	X1	X2
1	1	26	21.1
2	1	35	23.9
3	1	40	17.8
4	1	41	22.0
5	1	45	22.3
6	1	55	23.3
7	1	55	20.5
8	1	56	25.5
9	1	70	21.7
10	1	75	26.7
11	1	79	25.0
12	1	87	24.4
13	1	100	22.3
14	1	100	25.5
15	1	110	26.7
16	1	130	25.5
17	1	140	26.7

X- matrix (for fitting X_1 first).

(B)

FIREFLY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X, with intercept.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	1	194.42098784	194.42098784	3.08	0.0994
ERROR	15	945.46136510	63.03075767		
CORRECTED TOTAL	16	1139.88235294			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.170562	19.1986	7.93919125	41.35294118 $= \bar{y}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	194.42098784	3.08	0.0994

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	1	194.42098784	3.08	0.0994

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	$b_0 = 48.89620628$	10.39	0.0001	4.70688778
X1	$b_1 = -0.10308320$	-1.76	0.0994	0.05869385

compare with NOTES: p. 10.7.

OBSERVATION	OBSERVED VALUE y_i	PREDICTED VALUE $\hat{y}_i$	RESIDUAL $y_i - \hat{y}_i$
1	45.00000000	46.21604295	-1.21604295
2	40.00000000	45.28829411	-5.28829411
3	58.00000000	44.77287809	13.22712191
4	50.00000000	44.66979488	5.33020512
5	31.00000000	44.25746206	-13.25746206
6	52.00000000	43.22663002	8.77336998
7	54.00000000	43.22663002	10.77336998
8	38.00000000	43.12354681	-5.12354681
9	40.00000000	41.68038194	-1.68038194
10	28.00000000	41.16496592	-13.16496592
11	38.00000000	40.75263310	-2.75263310
12	36.00000000	39.92796746	-3.92796746
13	36.00000000	38.58788580	-2.58788580
14	46.00000000	38.58788580	7.41211420
15	40.00000000	37.55705376	2.44294624
16	31.00000000	35.49538966	-4.49538966
17	40.00000000	34.46455761	5.53544239

DEPENDENT VARIABLE: Y

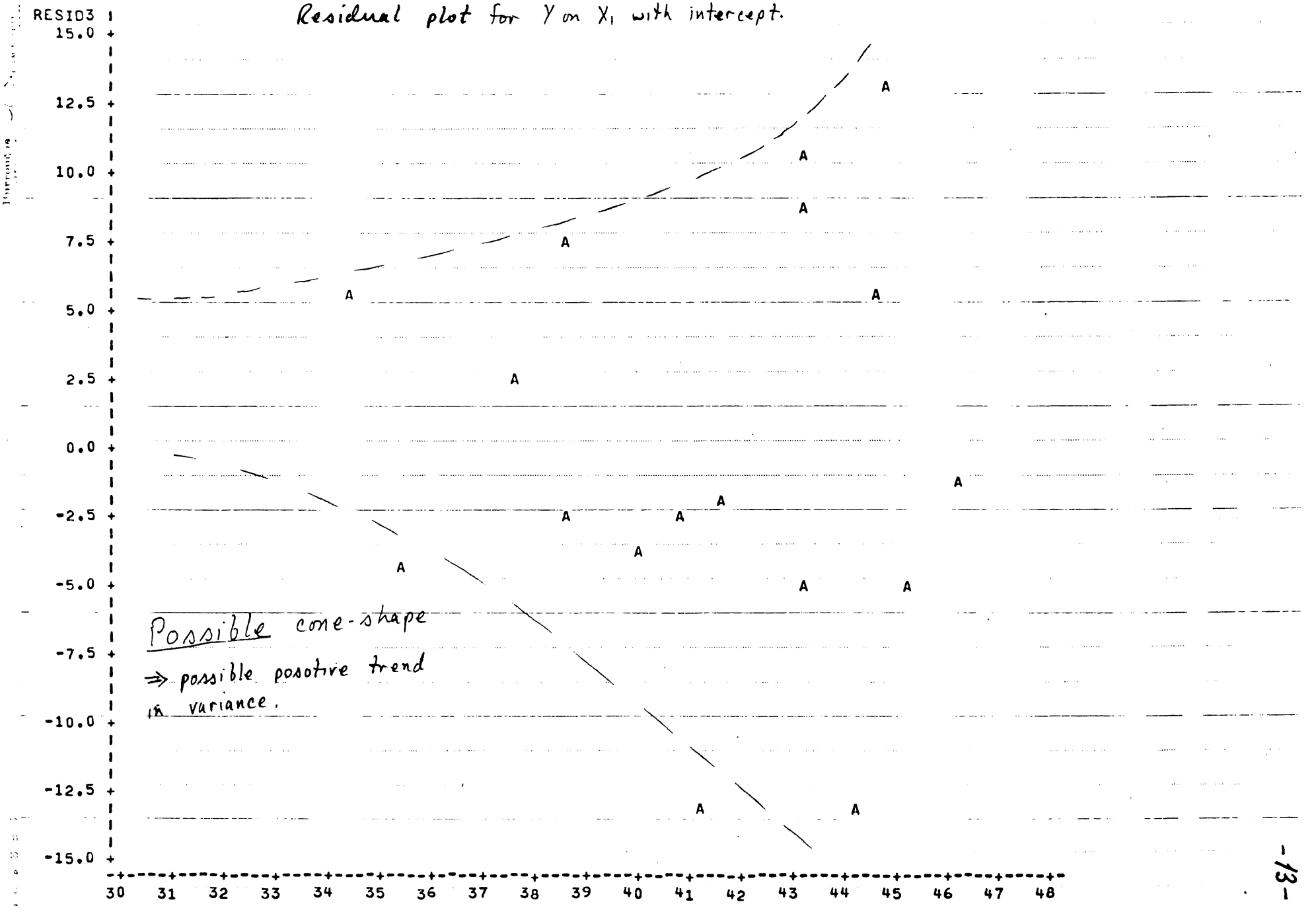
SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	945.46136510
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.09483671
DURBIN-WATSON D	2.15570071

(c)

FIREFLY DATA

PLOT OF RESID3*YHAT3 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Residual plot for Y on X_1 with intercept.



Possible cone-shape

⇒ possible posotive trend in variance.

(D)

FIRCFLY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y Regress Y on X_1 and X_2

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	456.73941145	228.36970573	4.68	0.0278
ERROR	14	683.14294149	48.79592439		
CORRECTED TOTAL	16	1139.88235294			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.400690	16.8922	6.98540796	41.35294118 = $\bar{y}$

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1 alone	1	194.42098784	3.98	0.0658
X2 after X_1	1	262.31842361	5.38	0.0361

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1 after X_2	1	0.00500314	0.00	0.9921
X2 after X_1	1	262.31842361	5.38	0.0361

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	b_0=91.47445073	4.86	0.0003	18.82510332
X1	b_1= 0.00069195	0.01	0.9921	0.06833915
X2	b_2=-2.12752930	-2.32	0.0361	0.91759899

compare with NOTES: pp. 10.10-12.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	45.00000000	46.60157419	-1.60157419
2	40.00000000	40.65072005	-0.65072005
3	58.00000000	53.63210873	4.36789127
4	50.00000000	44.69717766	5.30282234
5	31.00000000	44.06168682	-13.06168682
6	52.00000000	41.94107741	10.05892259
7	54.00000000	47.89815946	6.10184054
8	38.00000000	37.26120494	0.73879506
9	40.00000000	45.35550413	-5.35550413
10	28.00000000	34.72131757	-6.72131757
11	38.00000000	38.34080534	-0.34080534
12	36.00000000	39.62293883	-3.62293883
13	36.00000000	44.09974622	-8.09974622
14	46.00000000	37.29165246	8.70834754
15	40.00000000	34.74553719	5.25446281
16	31.00000000	37.31241213	-6.31241213
17	40.00000000	34.76629686	5.23370314

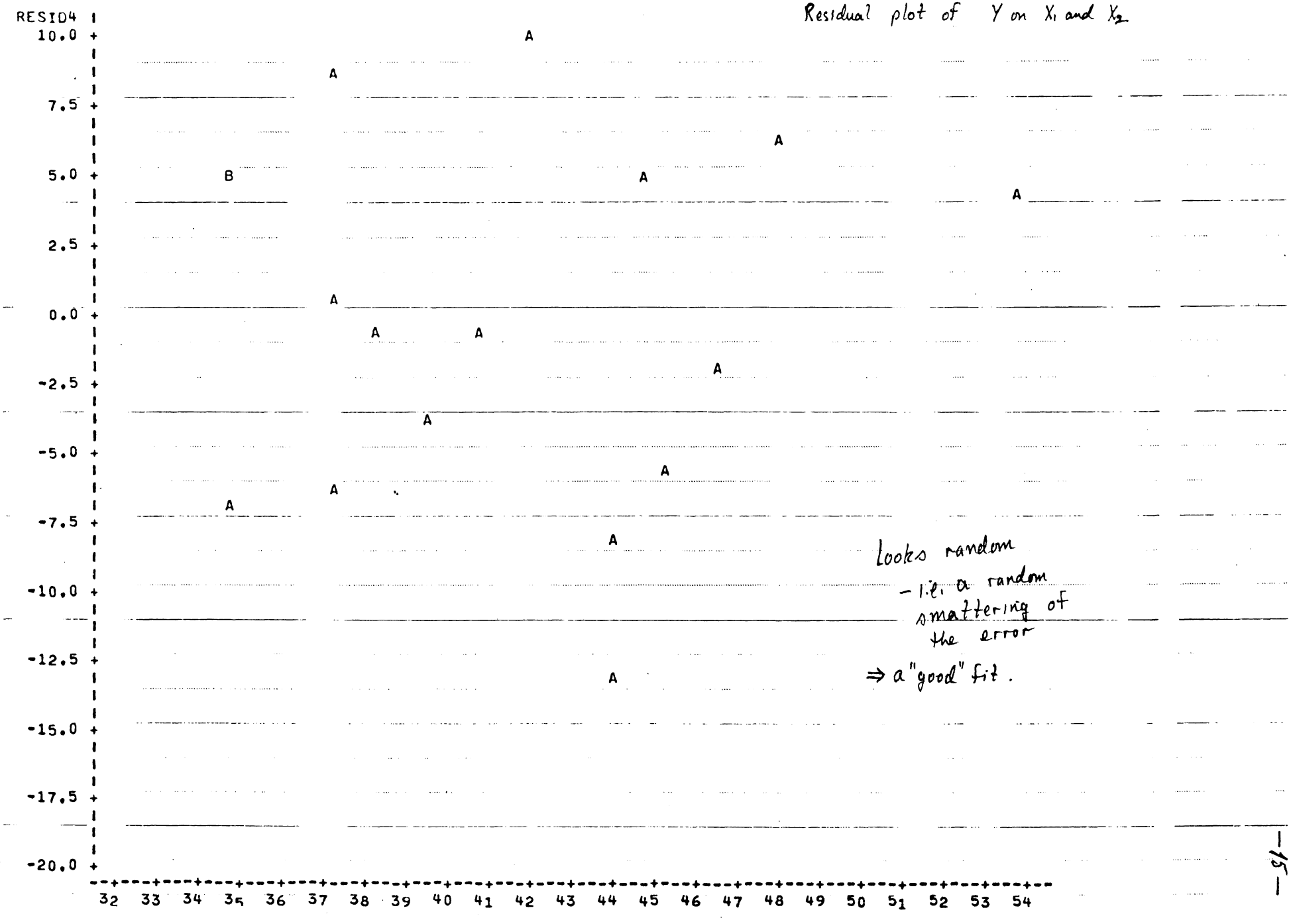
DEPENDENT VARIABLE: Y

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	683.14294149
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.20416682
DURBIN-WATSON D	2.36448237

(E)

FIREFLY DATA

PLOT OF RESID4*YHAT4 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot of Y on X_1 and X_2

Looks random
- i.e. a random smattering of the error

⇒ a "good" fit.

(F)

FIREFLY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X_2 and X_1

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	456.73941145	228.36970573	4.68	0.0278
ERROR	14	683.14294149	48.79592439		
CORRECTED TOTAL	16	1139.88235294			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.400690	16.8922	6.98540796	41.35294118

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X2 alone	1	456.73440831	9.36	0.0085
X1 after X_2	1	0.00500314	0.00	0.9921

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X2 after X_1	1	262.31842361	5.38	0.0361
X1 after X_2	1	0.00500314	0.00	0.9921

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	b_0 =91.47445073	4.86	0.0003	18.82510332
X2	b_2= -2.1275293	-2.32	0.0361	0.91759899
X1	b_1 = 0.00069199	0.01	0.9921	0.06833915

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	45.00000000	46.60157419	-1.60157419
2	40.00000000	40.65072005	-0.65072005
3	58.00000000	53.63210873	4.36789127
4	50.00000000	44.69717766	5.30282234
5	31.00000000	44.06168682	-13.06168682
6	52.00000000	41.94107741	10.05892259
7	54.00000000	47.89815946	6.10184054
8	38.00000000	37.26120494	0.73879506
9	40.00000000	45.35550413	-5.35550413
10	28.00000000	34.72131757	-6.72131757
11	38.00000000	38.34088534	-0.34088534
12	36.00000000	39.62293883	-3.62293883
13	36.00000000	44.09974622	-8.09974622
14	46.00000000	37.29165246	8.70834754
15	40.00000000	34.74553719	5.25446281
16	31.00000000	37.31241213	-6.31241213
17	40.00000000	34.76629686	5.23370314

Same as preceeding GLM procedure.

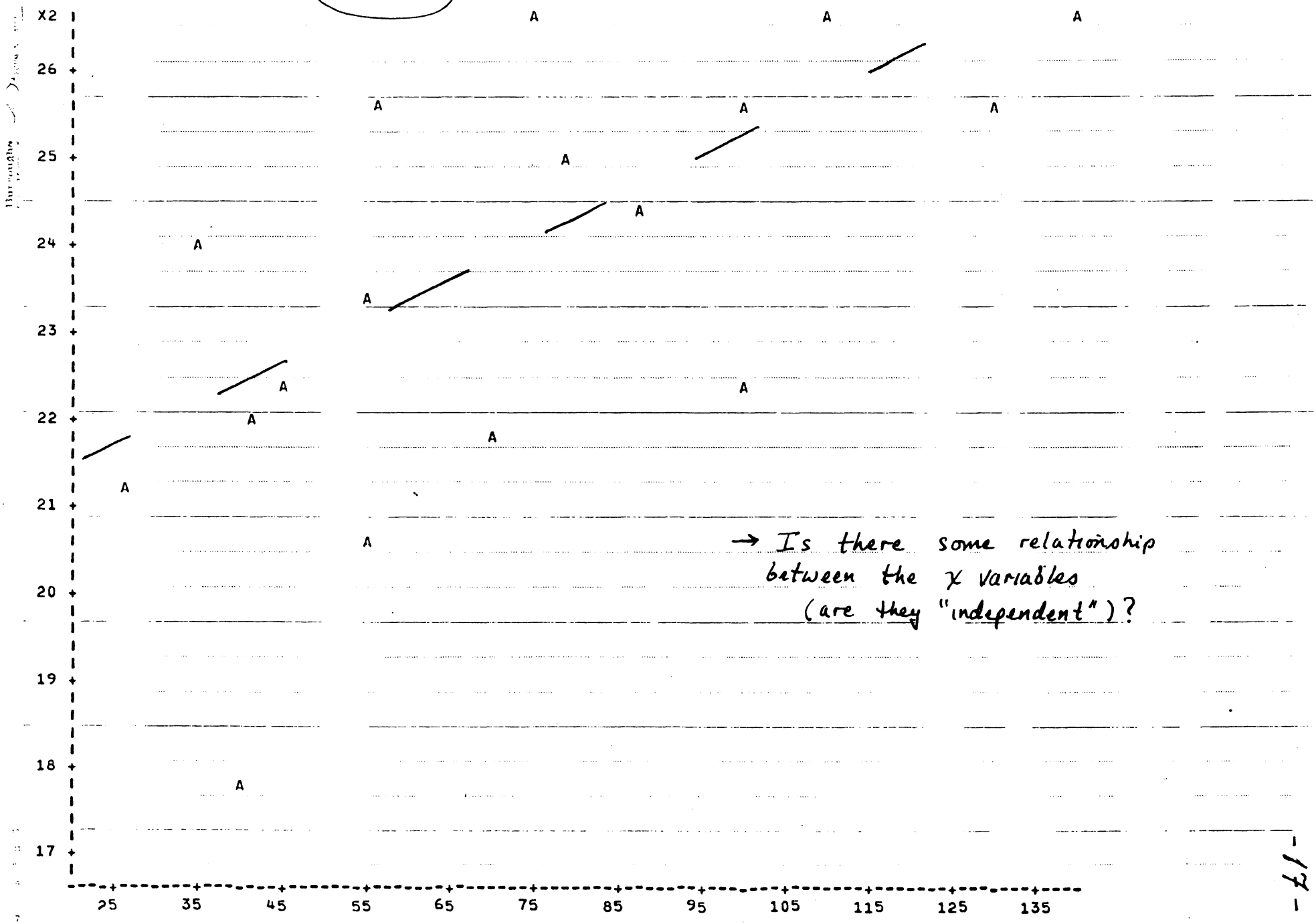
DEPENDENT VARIABLE: Y

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	683.14294149
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.20416682
DURBIN-WATSON D	2.36448237

(G)

FIREFLY DATA

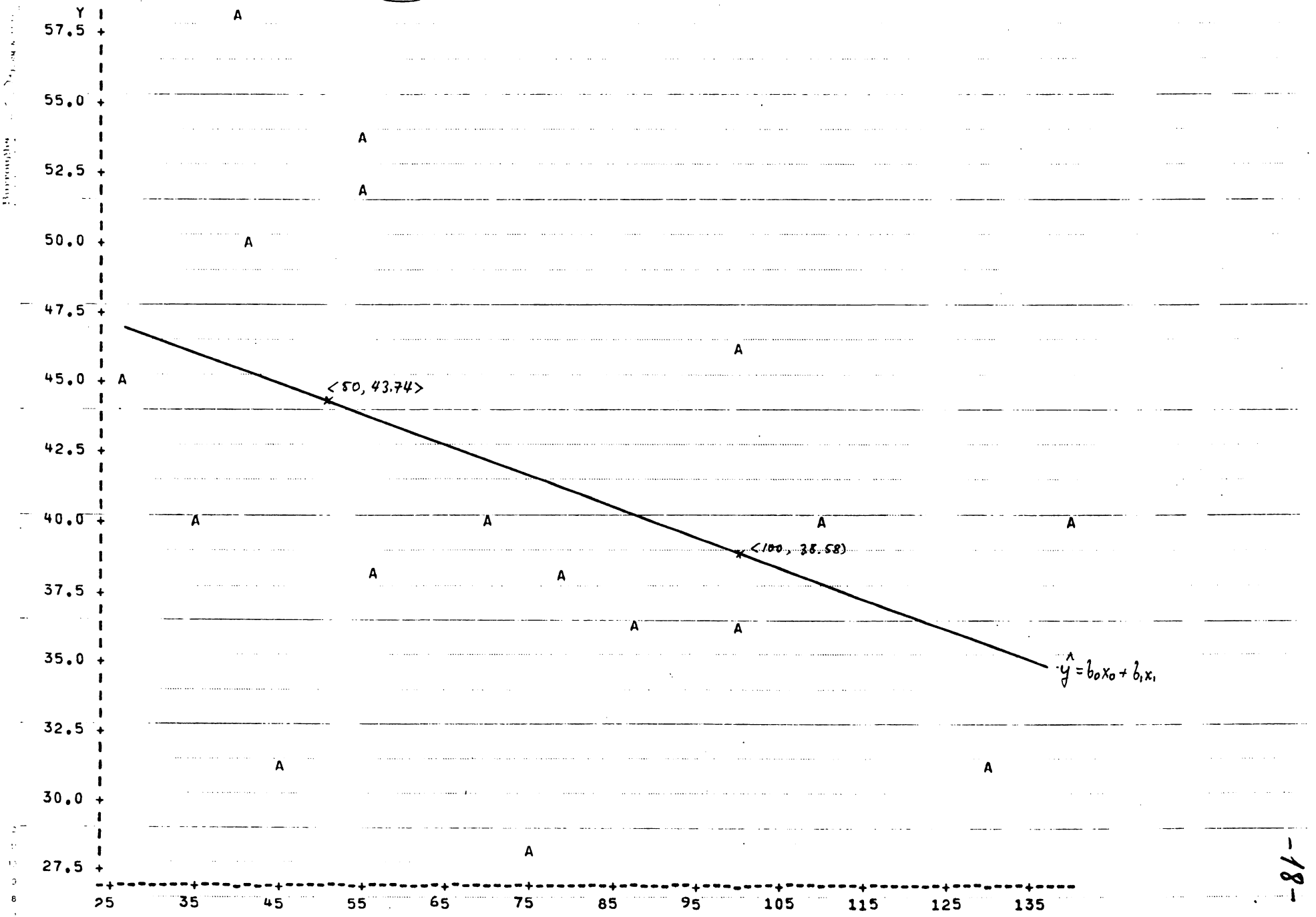
PLOT OF X2*X1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



→ Is there some relationship between the x variables (are they "independent")?

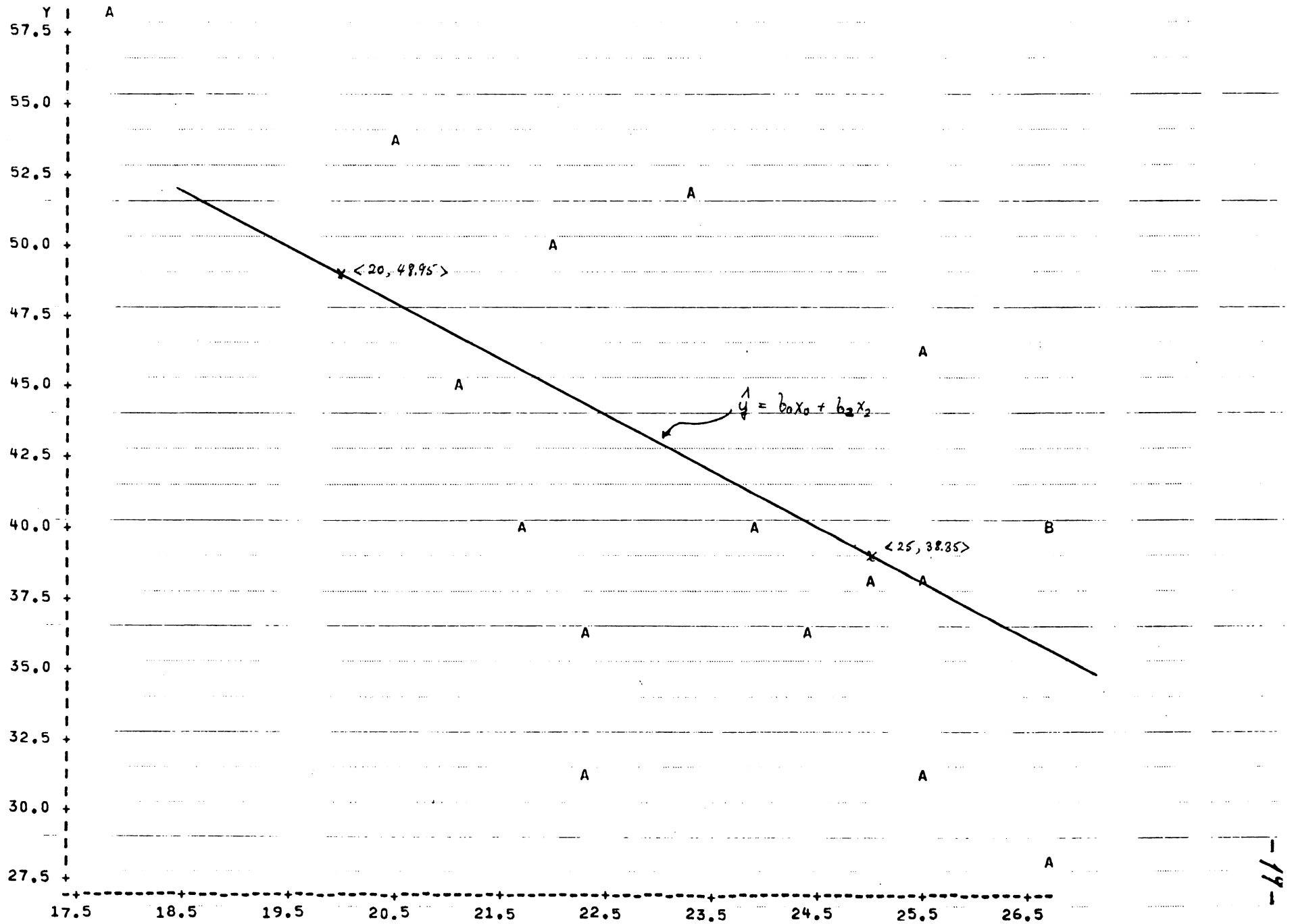
FIREFLY DATA

PLOT OF Y*X1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



FIREFLY DATA

PLOT OF Y*X2 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



(H)

FIREFLY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on X_2 with intercept.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	1	456.73440831	456.73440831	10.03	0.0064
ERROR	15	683.14794463	45.54319631		
CORRECTED TOTAL	16	1139.88235294			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.400686	16.3194	6.74856995	41.35294118

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X2	1	456.73440831	10.03	0.0064

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X2	1	456.73440831	10.03	0.0064

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	b_0 = 91.38158169	5.75	0.0001	15.88243590
X2	b_2 = -2.12144397	-3.17	0.0064	0.66990228

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	45.00000000	46.61911386	-1.61911386
2	40.00000000	40.67907074	-0.67907074
3	58.00000000	53.61987897	4.38012103
4	50.00000000	44.70981429	5.29018571
5	31.00000000	44.07338109	-13.07338109
6	52.00000000	41.95193712	10.04806288
7	54.00000000	47.89198025	6.10801975
8	38.00000000	37.28476038	0.71523962
9	40.00000000	45.34624748	-5.34624748
10	28.00000000	34.73902761	-6.73902761
11	38.00000000	38.34548237	-0.34548237
12	36.00000000	39.61834875	-3.61834875
13	36.00000000	44.07338109	-8.07338109
14	46.00000000	37.28476038	8.71523962
15	40.00000000	34.73902761	5.26097239
16	31.00000000	37.28476038	-6.28476038
17	40.00000000	34.73902761	5.26097239

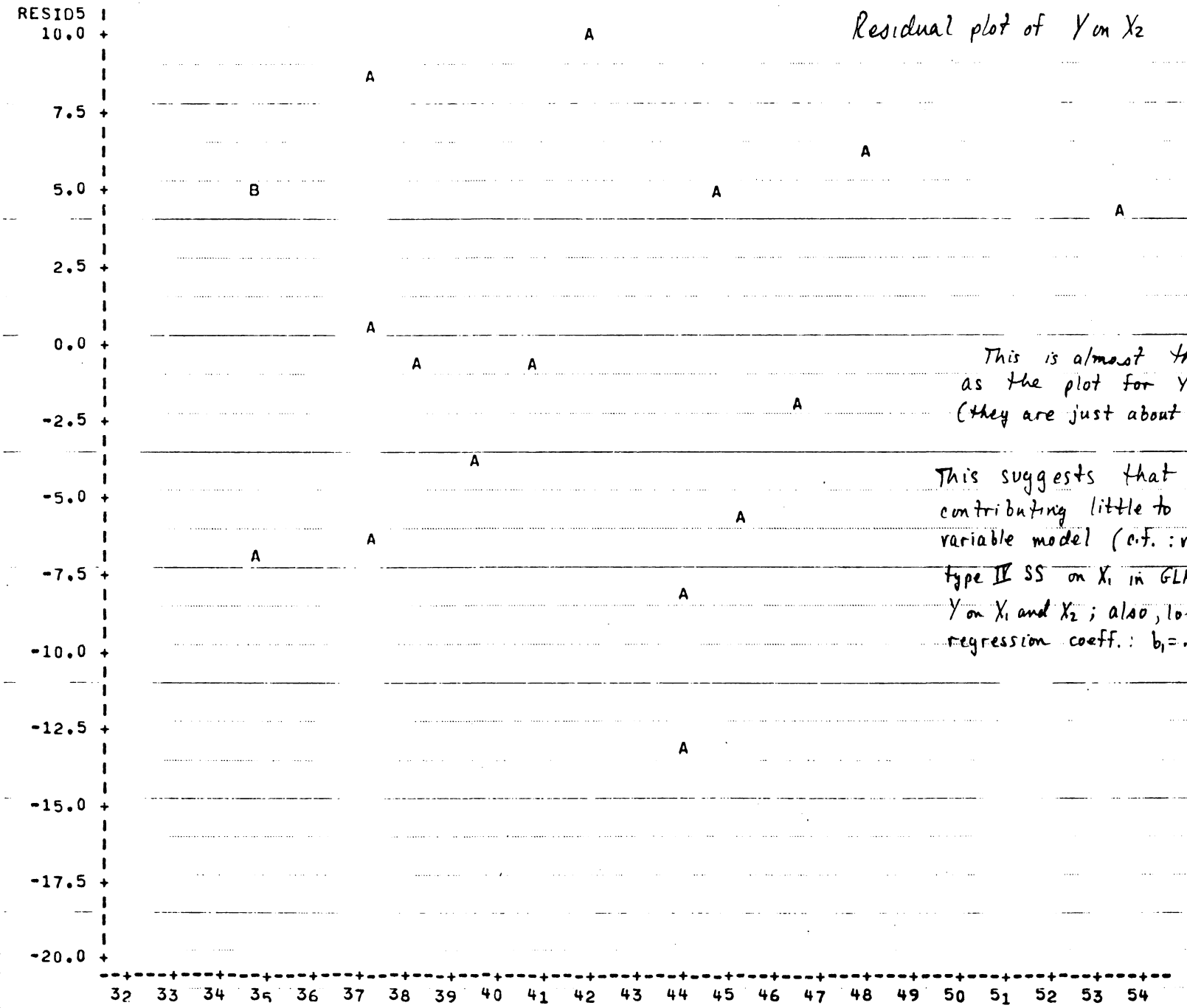
NOTE: these value are almost the same as those for Y on X_1 and X_2.

DEPENDENT VARIABLE: Y

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	683.14794463
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.20365578
DURBIN-WATSON D	2.36295900

FIREFLY DATA (J)

PLOT OF RESID5*YHAT5 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot of Y on X_2

This is almost the same as the plot for Y on X_1 and X_2 (they are just about indistinguishable).

This suggests that X_1 is contributing little to the two-variable model (c.f.: very low Type II SS on X_1 in GLM for Y on X_1 and X_2; also, low regression coeff.: $b_1 = .00069$).

1 ELECTRICITY LOAD DATA

NOTE: THE JOB VMPJ0001 HAS BEEN RUN UNDER RELEASE 79.28 OF SAS AT CORNELL UNIVER

```
1
2          DATA ELECLOAD;                    (unit 13)
3          TITLE ELECTRIC LOAD DATA;
4          INPUT DATE DAY $ TEMP Y;
5          X1=1;
6          X2=TEMP;
7          X3=0;
8          X4=0;
9          X5=0;
10         X6=0;
11         X7=TEMP;
12         IF DAY='SU' OR DATE=5 THEN LINK SUNDAY;
13         IF DAY='SA' THEN LINK SATURDAY;
14         RETURN;
15         SUNDAY: X5=1;
16         X6=TEMP;
17         X1=0;
18         X2=0;
19         RETURN;
20         SATURDAY: X3=1;
21         X4=TEMP;
22         X1=0;
23         X2=0;
24         RETURN;
25         CARDS;
```

creation of the x variables from the input variables

NOTE: DATA SET WORK.ELECLOAD HAS 31 OBSERVATIONS AND 11 VARIABLES. 141 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.29 SECONDS AND 116K.

```
57   (A)   PROC PRINT;
58         VAR X1 X3 X5 X7 X2 X4 X6;
```

print X-matrix, note no x_0

NOTE: THE PROCEDURE PRINT USED 0.20 SECONDS AND 116K AND PRINTED PAGE 1.

```
59         PROC GLM;
60   (B)   MODEL Y =X7/      P;
61         OUTPUT OUT=NEW1 RESIDUAL=RESID1 PREDICTED=YHAT1;
```

initial model: $y = \beta_0 x_0 + \beta_7 x_7 + \varepsilon$ page 13-10 in notes

NOTE: DATA SET WORK.NEW1 HAS 31 OBSERVATIONS AND 13 VARIABLES. 120 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.39 SECONDS AND 184K AND PRINTED PAGES 2 TO 3.

```
62         PROC PLOT;
63   (C)   PLOT Y*X7;
64   (D)   PLOT RESID1*YHAT1;
```

plot y versus x_7, same as Fig 13.2 in notes

plot residuals versus $\hat{y}$, contents of last 2 columns of Table 13.1 in notes.

NOTE: THE PROCEDURE PLOT USED 0.28 SECONDS AND 132K AND PRINTED PAGES 4 TO 5.

```
65         PROC GLM;
66   (E)   MODEL Y= X1 X3 X5/NOINT P;
67         OUTPUT OUT=NEW2 RESIDUAL=RESID2 PREDICTED=YHAT2;
```

in the sequence of models at the bottom of p. 13-14, this is the most reduced model, the 3 means model.

NOTE: DATA SET WORK.NEW2 HAS 31 OBSERVATIONS AND 15 VARIABLES. 105 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.41 SECONDS AND 188K AND PRINTED PAGES 6 TO 7.

```
68         PROC PLOT;
```

2

```
69  (F)  PLOT RESID2*YHAT2;
```
residual versus $\hat{y}$ plot

```
NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 124K AND PRINTED PAGE 8.

70       PROC GLM;
71  (G)  MODEL Y=X1 X3 X5 X7/NOINT P;
72       OUTPUT OUT=NEW3 RESIDUAL=RESID3 PREDICTED=YHAT3;
```
no common intercept
3 different intercepts
+ 1 common slope
the 'middle' reduced model

```
NOTE: DATA SET WORK.NEW3 HAS 31 OBSERVATIONS AND 17 VARIABLES. 93 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.43 SECONDS AND 188K AND PRINTED PAGES 9 TO 10.

73  (H)  PROC PLOT;
74       PLOT RESID3*YHAT3;
```
residual plot for the above model, can be written $\hat{y} = \bar{y}_w x_1 + \bar{y}_{sa} x_3 + \bar{y}_{su} x_5 + b_2(x_2-\bar{x}_2) + b_4(x_4-\bar{x}_4) + b_6(x_6-\bar{x}_6)$

w = weekday sa = saturday su = sunday

```
NOTE: THE PROCEDURE PLOT USED 0.20 SECONDS AND 124K AND PRINTED PAGE 11.

75       PROC GLM;
76  (I)  MODEL Y= X1 X3 X5 X7 X2 X4 X6/NOINT P;
77       OUTPUT OUT=NEW4 RESIDUAL=RESID4 PREDICTED=YHAT4;
```
full model:
3 different intercepts
+ 3 different slopes.
note order of variables

```
NOTE: DATA SET WORK.NEW4 HAS 31 OBSERVATIONS AND 19 VARIABLES. 83 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.44 SECONDS AND 188K AND PRINTED PAGES 12 TO 14.

78  (J)  PROC PLOT;
79       PLOT RESID4*YHAT4;
```
residual plot for full model

$\hat{y} = \bar{y}_w x_1 + \bar{y}_{sa} x_3 + \bar{y}_{su} x_5 + b_2(x_2-\bar{x}_2) + b_4(x_4-\bar{x}_4) + b_6(x_6-\bar{x}_6)$

```
NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 126K AND PRINTED PAGE 15.
```

ELECTRIC LOAD DATA

(A)

OBS	X1	X3	X5	X7	X2	X4	X6
1	1	0	0	77	77	0	0
2	1	0	0	75	75	0	0
3	0	1	0	74	0	74	0
4	0	0	1	78	0	0	78
5	0	0	1	77	0	0	77
6	1	0	0	72	72	0	0
7	1	0	0	79	79	0	0
8	1	0	0	81	81	0	0
9	1	0	0	85	85	0	0
10	0	1	0	83	0	83	0
11	0	0	1	82	0	0	82
12	1	0	0	75	75	0	0
13	1	0	0	77	77	0	0
14	1	0	0	75	75	0	0
15	1	0	0	77	77	0	0
16	1	0	0	77	77	0	0
17	0	1	0	82	0	82	0
18	0	0	1	74	0	0	74
19	1	0	0	72	72	0	0
20	1	0	0	71	71	0	0
21	1	0	0	73	73	0	0
22	1	0	0	78	78	0	0
23	1	0	0	78	78	0	0
24	0	1	0	74	0	74	0
25	0	0	1	77	0	0	77
26	1	0	0	79	79	0	0
27	1	0	0	73	73	0	0
28	1	0	0	75	75	0	0
29	1	0	0	69	69	0	0
30	1	0	0	71	71	0	0
31	0	1	0	68	0	68	0

X - matrix

compare with table on page 13-13.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	1	32474.31459229	32474.31459229	2.46	0.1277
ERROR	29	383068.45960126	13209.25722763		
CORRECTED TOTAL	30	415542.77419355			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.078149	10.9180	114.93153278	1052.67741935

≈ 7.8% should be the same as the R^2 in the notes p. 13-10 states as 78%

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X7	1	32474.31459229	2.46	0.1277

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X7	1	32474.31459229	2.46	0.1277

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	434.62329486	1.10	0.2799	394.72114585
X7	8.12539349	1.57	0.1277	5.18219325

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	1156.00000000	1060.27859391	95.72140609
2	1073.00000000	1044.02780693	28.97219307
3	893.00000000	1035.90241343	-142.90241343
4	818.00000000	1068.40398741	-250.40398741
5	889.00000000	1060.27859391	-171.27859391
6	1045.00000000	1019.65162644	25.34837356
7	1155.00000000	1076.52938090	78.47061910
8	1213.00000000	1092.78016789	120.21983211
9	1273.00000000	1125.28174187	147.71825813
10	1092.00000000	1109.03095488	-17.03095488
11	899.00000000	1100.90556139	-201.90556139
12	1130.00000000	1044.02780693	85.97219307
13	1136.00000000	1060.27859391	75.72140609
14	1111.00000000	1044.02780693	66.97219307
15	1121.00000000	1060.27859391	60.72140609
16	1174.00000000	1060.27859391	113.72140609
17	1034.00000000	1100.90556139	-66.90556139
18	866.00000000	1035.90241343	-169.90241343
19	1105.00000000	1019.65162644	85.34837356
20	1066.00000000	1011.52623295	54.47376705

← last 3 columns of table 13.1

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

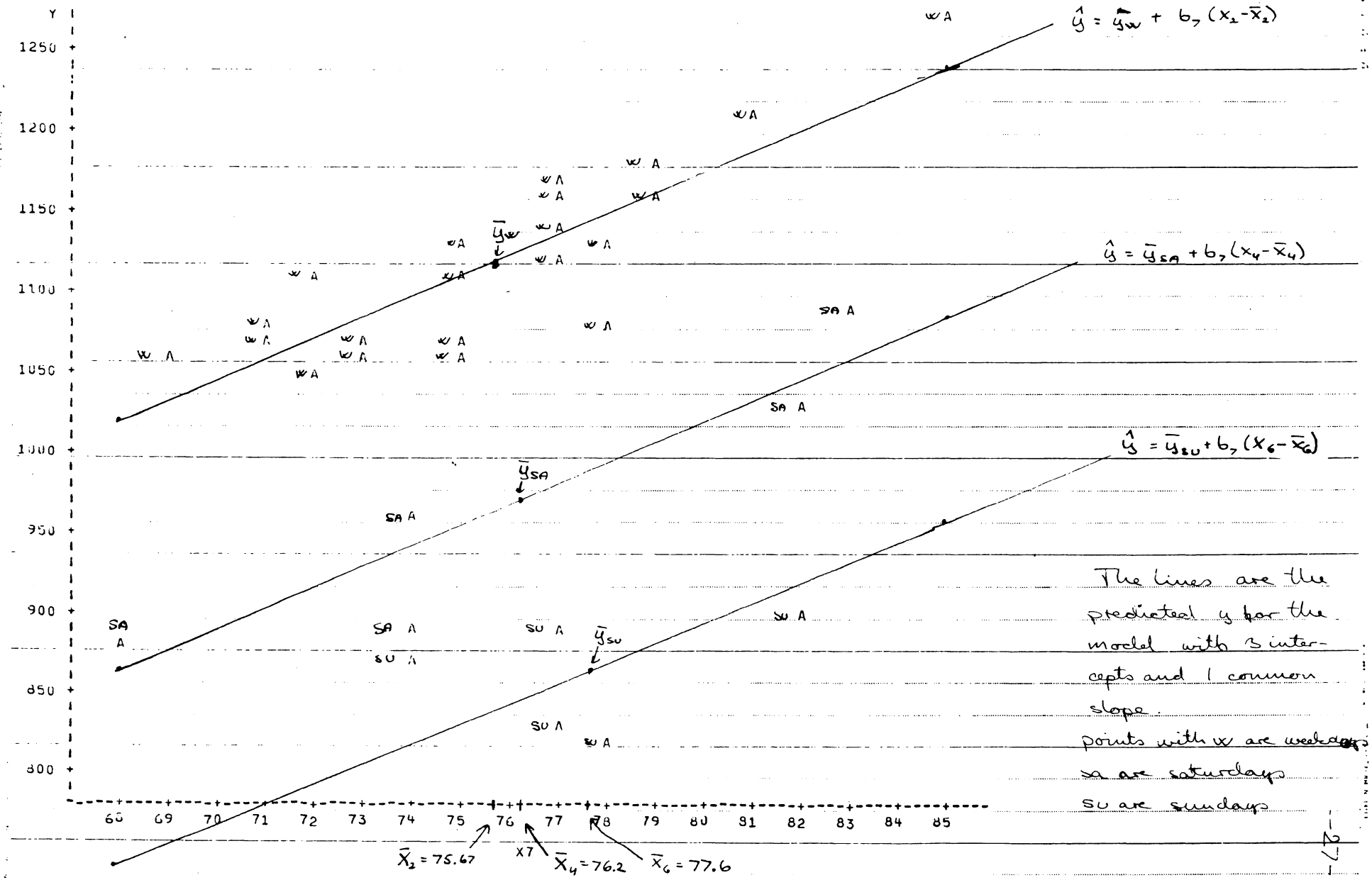
OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
21	1062.00000000	1027.77701994	34.22298006
22	1078.00000000	1068.40398741	9.59601259
23	1126.00000000	1068.40398741	57.59601259
24	961.00000000	1035.90241343	-74.90241343
25	825.00000000	1060.27859391	-235.27859391
26	1179.00000000	1076.52938090	102.47061910
27	1068.00000000	1027.77701994	40.22298006
28	1063.00000000	1044.02780693	18.97219307
29	1061.00000000	995.27544596	65.72455404
30	1081.00000000	1011.52623295	69.47376705
31	880.00000000	987.15005247	-107.15005247

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	383068.45960126
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	0.25418122
DURBIN-WATSON D	1.43774713

second part of table 13.1.

ELECTRIC LOAD DATA

PLOT OF Y*X7 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

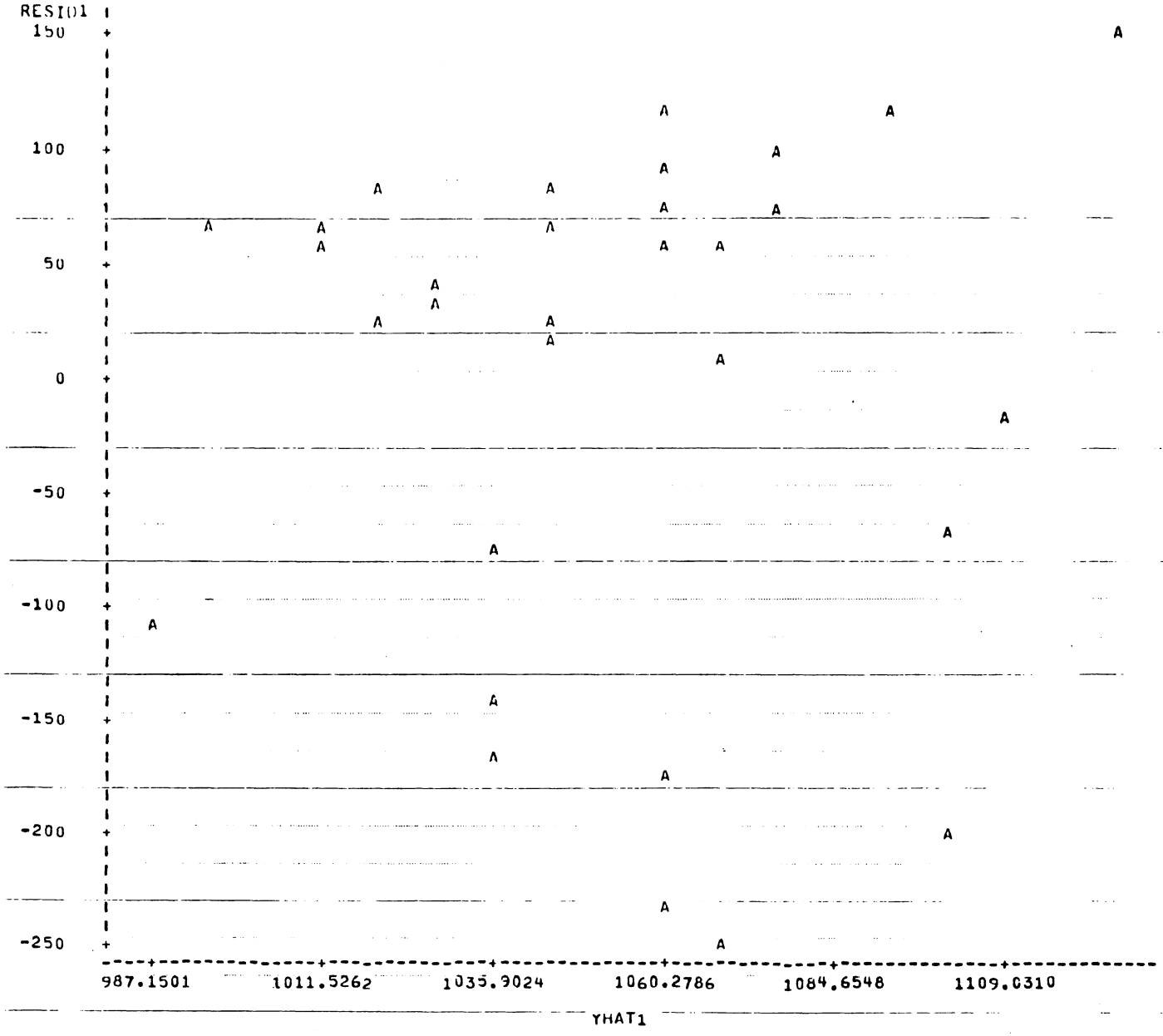


The lines are the predicted $\hat{y}$ for the model with 3 intercepts and 1 common slope.
points with W are weekdays
SA are saturdays
SU are sundays

(D)

ELECTRIC LOAD DATA

PLOT OF RESID1*YHAT1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot for 1 intercept and 1 slope

Contents of Table 13.1

(17)

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	3	34660693.99047619	11553564.66349206	3027.01	0.0001
ERROR	28	106871.00952381	3816.82176871		
UNCORRECTED TOTAL	31	34767565.00000000			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.996926	5.8689	61.78043192	1052.67741935

Seq. SS

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	26243952.19047619	6875.66	0.0001
X3 after X1	1	4723920.80000000	1237.66	0.0001
X5 after X1 & X3	1	3692841.80000000	967.52	0.0001

Part. SS

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1 after X3 & X5	1	26243952.19047619	6875.66	0.0001
X3 after X1 & X5	1	4723920.80000000	1237.66	0.0001
X5 after X1 & X3	1	3692841.80000000	967.52	0.0001

note that sequential and partial SS are the same, when fitting 3 different means

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	b_1 = 1117.90476190 = $\bar{y}_{1\cdot}$	82.92	0.0001	13.48159551
X3	b_3 = 972.00000000 = $\bar{y}_{3\cdot}$ - $\bar{y}_{5\cdot}$	35.18	0.0001	27.62904909
X5	b_5 = 859.40000000 = $\bar{y}_{5\cdot}$	31.10	0.0001	27.62904909

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	1156.00000000	1117.90476190	38.09523810
2	1073.00000000	1117.90476190	-44.90476190
3	893.00000000	972.00000000	-79.00000000
4	818.00000000	859.40000000	-41.40000000
5	889.00000000	859.40000000	29.60000000
6	1045.00000000	1117.90476190	-72.90476190
7	1155.00000000	1117.90476190	37.09523810
8	1213.00000000	1117.90476190	95.09523810
9	1273.00000000	1117.90476190	155.09523810
10	1092.00000000	972.00000000	120.00000000
11	899.00000000	859.40000000	39.60000000
12	1130.00000000	1117.90476190	12.09523810
13	1136.00000000	1117.90476190	18.09523810
14	1111.00000000	1117.90476190	-6.90476190

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

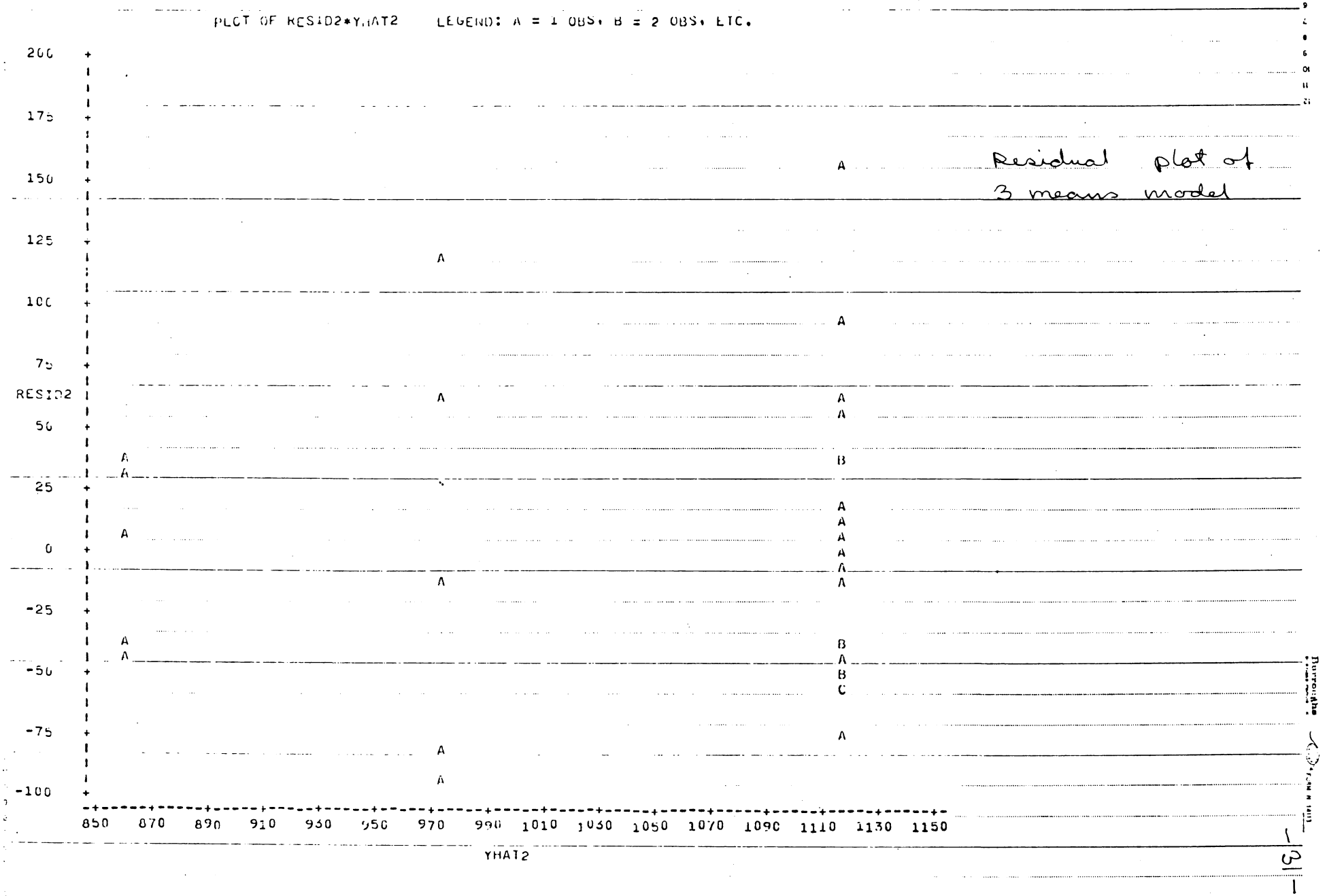
OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
15	1121.00000000	1117.90476190	3.09523810
16	1174.00000000	1117.90476190	56.09523810
17	1034.00000000	972.00000000	62.00000000
18	866.00000000	859.40000000	6.60000000
19	1105.00000000	1117.90476190	-12.90476190
20	1066.00000000	1117.90476190	-51.90476190
21	1062.00000000	1117.90476190	-55.90476190
22	1078.00000000	1117.90476190	-39.90476190
23	1126.00000000	1117.90476190	8.09523810
24	961.00000000	972.00000000	-11.00000000
25	825.00000000	859.40000000	-34.40000000
26	1179.00000000	1117.90476190	61.09523810
27	1068.00000000	1117.90476190	-49.90476190
28	1063.00000000	1117.90476190	-54.90476190
29	1061.00000000	1117.90476190	-56.90476190
30	1081.00000000	1117.90476190	-36.90476190
31	880.00000000	972.00000000	-92.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	106871.00952381
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.53489004
DURBIN-WATSON D	0.83744222

always 0 when fitting 1 or more means or 1 or more intercepts in a model, No intercepts or means will make it different from 0.

(F)

ELECTRIC LOAD DATA



Residual plot of 3 means model

(G)

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	4	34738249.28002397	8684562.32000599	7998.55	0.0001
ERROR	27	29315.71997603	1085.76740652		
UNCORRECTED TOTAL	31	34767565.00000000			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.999157	3.1302	32.95098491	1052.67741935

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	R(1)=26243932.19047619	24170.66	0.0001
X3 after X1	1	R(3|1)=4723920.00000000	4350.77	0.0001
X5 after X1 X3	1	R(5|1,3)=3692841.80000000	3401.14	0.0001
X7 after X1 X3 X5	1	R(7|1,3,5)=77555.28954779	71.43	0.0001

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1 after X3 X5 X7	1	R(1|3,5,7)=1934.52471220	1.78	0.1931
X3 after X1 X5 X7	1	R(3|1,5,7)=0.00006609	0.00	0.9998
X5 after X1 X3 X7	1	R(5|1,3,7)=1325.61070453	1.22	0.2709
X7 after X1 X3 X5	1	R(7|1,3,5)=77555.28954779	71.43	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	152.73674326 = b_1	1.33	0.1931	114.42601814
X3	0.02903497 = b_3	0.00	0.9998	115.94506590
X5	-130.42869930 = b_5	-1.10	0.2709	118.04118545
X7	12.75552448 = b_7	8.45	0.0001	1.50924939

Note b_1, b_3 & b_5 are not the means in this model

These b's are the estimates for the reduced model that seems most adequate.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	1156.00000000	1134.91212787	21.08787213
2	1073.00000000	1109.40107892	-36.40107892
3	893.00000000	943.93784615	-50.93784615
4	818.00000000	864.50220979	-46.50220979
5	889.00000000	851.74668531	37.25331469
6	1045.00000000	1071.13450549	-26.13450549
7	1155.00000000	1160.42317682	-5.42317682
8	1213.00000000	1185.93422577	27.06577423
9	1273.00000000	1236.95632368	36.04367632
10	1092.00000000	1058.73756643	33.26243357
11	899.00000000	915.52430769	-16.52430769

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

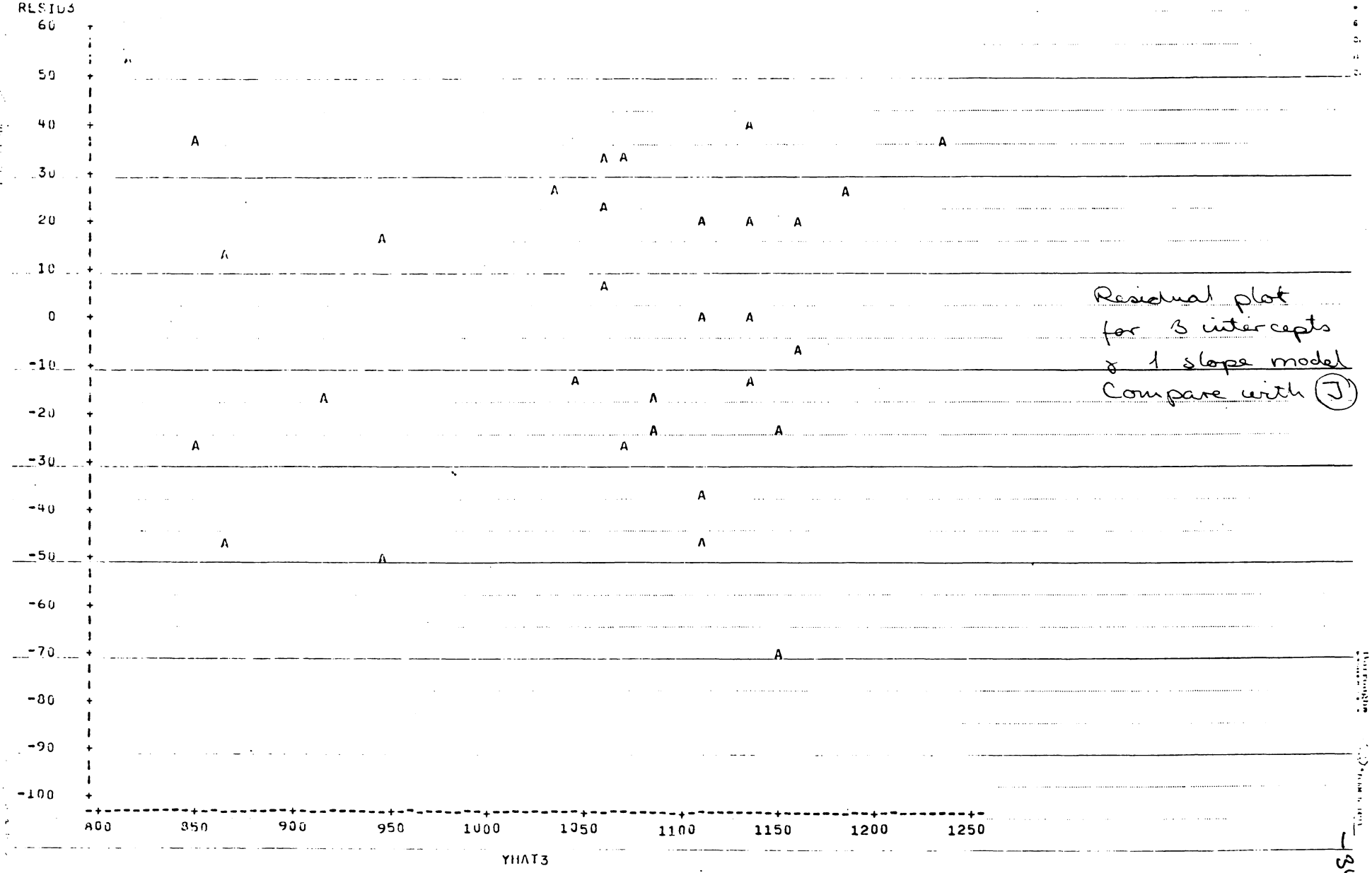
OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
12	1130.00000000	1109.40107892	20.59892108
13	1136.00000000	1134.91212787	1.08787213
14	1111.00000000	1109.40107892	1.59892108
15	1121.00000000	1134.91212787	-13.91212787
16	1174.00000000	1134.91212787	39.08787213
17	1034.00000000	1045.98204196	-11.98204196
18	865.00000000	813.48011189	52.51988811
19	1105.00000000	1071.33450549	33.66549451
20	1066.00000000	1058.37898102	7.62101898
21	1062.00000000	1083.89002997	-21.89002997
22	1078.00000000	1147.66765235	-69.66765235
23	1126.00000000	1147.66765235	-21.66765235
24	961.00000000	943.93784615	17.06215385
25	825.00000000	851.74668531	-26.74668531
26	1179.00000000	1160.42317682	18.57682318
27	1068.00000000	1083.89002997	-15.89002997
28	1063.00000000	1109.40107892	-46.40107892
29	1061.00000000	1032.86793207	28.13206793
30	1081.00000000	1058.37898102	22.62101898
31	880.00000000	867.40469930	12.59530070

SUM OF RESIDUALS	-0.00000000
SUM OF SQUARED RESIDUALS	29515.71997602
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.13742647
DURBIN-WATSON D	1.70456629

(H)

ELECTRIC LOAD DATA

PLOT OF RESID3*YHAT3 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot for 3 intercepts & 1 slope model
Compare with (J)

(I)

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

The model has 7 parameters but 1 is superfluous therefore the model has 6 df = 3 for intercepts + 3 for slopes

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	6	34740917.03721010	5790152.84953502	5432.09	0.0001
ERROR	25	26647.96278990	1065.91611160		
UNCORRECTED TOTAL	31	34767565.00000000			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.999234	3.1015	32.64857075	1052.07741935

the ANOVA table on p. 13-15 can be completed using these numbers:

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	26243932.19047619	24621.01	0.0001
X3 after X1	1	4723920.00000000	4431.79	0.0001
X5 after X1 & X3	1	3692841.80000000	3464.48	0.0001
X7 after X1, X3 & X5	1	77555.29354779	72.76	0.0001
X2 after X1, X3, X5 & X7	1	220.95125542	0.21	0.6531
X4 after X1, X3, X5, X7 & X2	1	2447.26592771	2.30	0.1423
X6 after X1, X3, X5, X7, X2, X4	0	0.00000000	.	.

sum of these = R(1,3,5) = SS for 3 different means

= R(7|1,3,5) = SS for 1 common slope

sum of these = R(2,4,6|1,3,5,7) = SS for different slopes

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1 after the other x's	1	613.33455571	0.58	0.4551
X3 " " "	1	103.72400562	0.10	0.7577
X5 " " "	1	1603.34116067	1.50	0.2314
X7 " " "	0	0.00000000	.	.
X2 " " "	0	0.00000000	.	.
X4 " " "	0	0.00000000	.	.
X6 " " "	0	0.00000000	.	.

1 of the slopes is redundant

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1	110.83421927	0.76	0.4551	146.08130199
X3	-62.14285714	-0.31	0.7577	199.21089667
X5	539.65060241	1.23	0.2314	439.94005229
X7	4.12048193 B	0.73	0.4739	5.66620746
X2	9.17882040 B	1.54	0.1373	5.98533366
X4	9.45094664 B	1.52	0.1423	6.23729439
X6	0.00000000 B	.	.	.

← the std error here is 3½ times bigger than in the reduced model with 1 slope, and it's bigger than b_7 itself.

↖ X_2 or X_4 or X_6 is superfluous when X_7 is fitted first

Since you have redundant variables (1) you get the note on the following page from SAS:

ELECTRIC LOAD DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: B(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	1156.00000000	1135.65049834	20.34950166
2	1073.00000000	1109.03189369	-36.03189369
3	893.00000000	942.14285714	-49.14285714
4	818.00000000	861.04819277	-43.04819277
5	889.00000000	856.92771084	32.07228916
6	1045.00000000	1069.10398671	-24.10398671
7	1155.00000000	1162.26910299	-7.26910299
8	1213.00000000	1188.88770764	24.11229236
9	1273.00000000	1242.12491694	30.87508306
10	1092.00000000	1064.28571429	27.71428571
11	899.00000000	877.53012048	21.46987952
12	1130.00000000	1109.03189369	20.96810631
13	1136.00000000	1135.65049834	0.34950166
14	1111.00000000	1109.03189369	1.96810631
15	1121.00000000	1135.65049834	-14.65049834
16	1174.00000000	1135.65049834	38.34950166
17	1034.00000000	1050.71428571	-16.71428571
18	866.00000000	844.56626506	21.43373494
19	1105.00000000	1069.10398671	35.89601329
20	1066.00000000	1055.79468439	10.20531561
21	1062.00000000	1082.41328904	-20.41328904
22	1078.00000000	1148.95980066	-70.95980066
23	1126.00000000	1148.95980066	-22.95980066
24	961.00000000	944.14285714	16.85714286
25	825.00000000	856.92771084	-31.92771084
26	1179.00000000	1162.26910299	16.73089701
27	1068.00000000	1082.41328904	-14.41328904
28	1063.00000000	1109.03189369	-46.03189369
29	1061.00000000	1029.17607973	31.82392027
30	1081.00000000	1055.79468439	25.20531561
31	880.00000000	860.71428571	19.28571429

ELECTRIC LOAD DATA

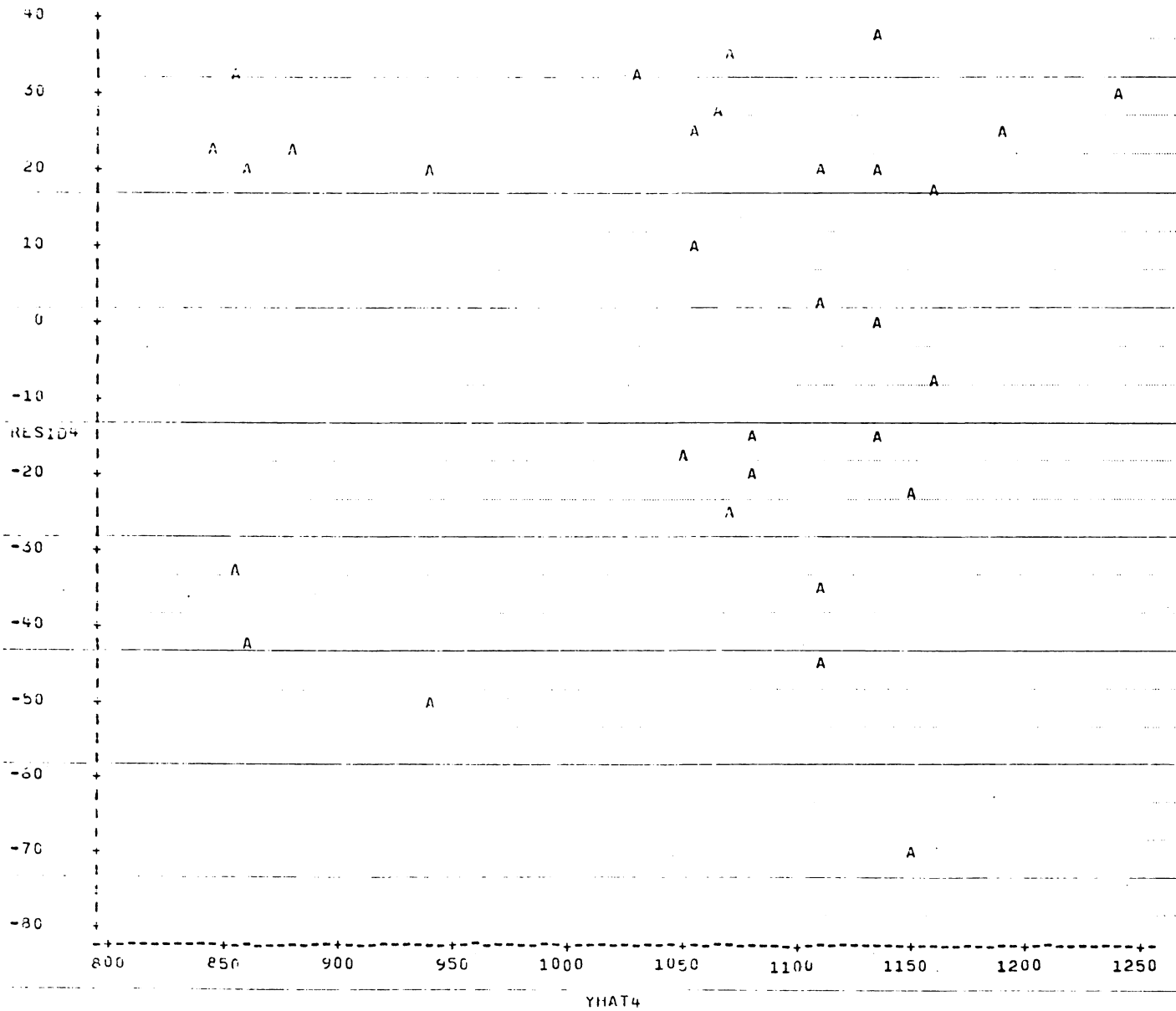
GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	26647.90272990
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.17803463
DURBIN-WATSON D	1.61443345

ELECTRIC LOAD DATA

PLOT OF RESID4*YHAT4 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot for the full model with 3 means and 3 slopes.

The plot looks a lot like the residual plot for 3 means and 1 slope on page 34, both with respect to pattern and to magnitude.

LEAFHOPPER DATA

Objectives: (i) Fit general means model.
(ii) Calculate residual mean squares and means.
(iii) Estimate contrasts and associated std. errors.

Unit 16

```
2          DATA LEAFHOPPER;
3               INPUT TRTS $ DAYS;
4               X0=1;
5               IF TRTS='CONTROL' THEN X1=1;
6                    ELSE X1=0;
7               IF TRTS='SUCROSE' THEN X2=1;
8                    ELSE X2=0;
9               IF TRTS='GLUCOSE' THEN X3=1;
10                   ELSE X3=0;
11              IF TRTS='FRUCTOSE' THEN X4=1;
12                   ELSE X4=0;
13         CARDS;

NOTE: DATA SET WORK.LEAFHOPPER HAS 8 OBSERVATIONS AND 7 VARIABLES. 217 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.21 SECONDS AND 116K.

22         TITLE LEAFHOPPER DATA;
23         PROC GLM;
24              CLASS TRTS;
25              MODEL DAYS=TRTS/NOINT SOLUTION P;
26              OUTPUT OUT=NEW6 RESIDUAL=RESID6   PREDICTED=YHAT6;
27         ESTIMATE 'CONTROL VS SUGARS'
28              TRTS 1 -.333333 -.333333 -.333333/E;
29         ESTIMATE '6-CARBONS VS SUCROSE'
30              TRTS 0 -.5 -.5 1/E;
31         ESTIMATE 'FRUCTOSE VS GLUCOSE'
32              TRTS 0 -1 1 0/E;
33              MEANS TRTS;

NOTE: DATA SET WORK.NEW6 HAS 8 OBSERVATIONS AND 9 VARIABLES. 171 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.44 SECONDS AND 190K AND PRINTED PAGES 1 TO 7.

34         PROC PLOT;
35              PLOT RESID6*YHAT6;

NOTE: THE PROCEDURE PLOT USED 0.24 SECONDS AND 124K AND PRINTED PAGE 8.

36         PROC PRINT;
37              VAR X0-X4;

NOTE: THE PROCEDURE PRINT USED 0.16 SECONDS AND 116K AND PRINTED PAGE 9.

38         PROC GLM;
39              MODEL DAYS=X1-X4/P;
40

NOTE: DATA SET WORK.NEW7A HAS 8 OBSERVATIONS AND 11 VARIABLES. 141 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.30 SECONDS AND 186K AND PRINTED PAGES 10 TO 11.
```

Group TRTS variable for fitting using CLASS statement.

Regress DAYS on TRTS with no intercept (need SOLUTION statement after CLASS statement to print parameter estimates), for the general means model.

Estimate different contrasts.

Calculate TRTS means.

Plot $(y_i - \hat{y}_i)$ vs. $\hat{y}_i$

Print X matrix for equal means / general means model sequence.

Regress DAYS on X1, X2, X3, X4 with intercept for general means model (don't need residual plot, residuals are the same as in the general means model).

LEAFHOPPER DATA

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRTS	4	CONTROL FRUCTOSE GLUCOSE SUCROSE

NUMBER OF OBSERVATIONS IN DATA SET = 8

Summary of CLASS statement results

DEPENDENT VARIABLE: DAYS

ESTIMABLE FUNCTIONS FOR CONTROL VS SUGARS

1st contrast vector

EFFECT		COEFFICIENTS
TRTS	CONTROL	1
	FRUCTOSE	-0.333333
	GLUCOSE	-0.333333
	SUCROSE	-0.333333

DEPENDENT VARIABLE: DAYS

ESTIMABLE FUNCTIONS FOR 6-CARBONS VS SUCROSE

2nd contrast vector

EFFECT		COEFFICIENTS
TRTS	CONTROL	0
	FRUCTOSE	-0.5
	GLUCOSE	-0.5
	SUCROSE	1

DEPENDENT VARIABLE: DAYS

ESTIMABLE FUNCTIONS FOR FRUCTOSE VS GLUCOSE

3rd contrast vector

EFFECT		COEFFICIENTS
TRTS	CONTROL	0
	FRUCTOSE	-1
	GLUCOSE	1
	SUCROSE	0

MEANS

TRTS	N	DAYS
CONTROL	2	2.00000000
FRUCTOSE	2	2.20000000
GLUCOSE	2	2.90000000
SUCROSE	2	3.80000000

Calculate treatment means.
cf. NOTES: p. 16.8.

LEAFHOPPER DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: DAYS Regress DAYS on TRTS with no intercept

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	4	63.38000000	15.84500000	211.27	0.0001
ERROR	4	0.30000000	0.07500000		
UNCORRECTED TOTAL	8	63.68000000			

cf. NOTES: p.16.2.

R-SQUARE	C.V.	STD DEV	DAYS MEAN
0.995289	10.0500	0.27386128	2.72500000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
TRTS	4	63.38000000	211.27	0.0001

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
TRTS	4	63.38000000	211.27	0.0001

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
TRTS	CONTROL	2.00000000	10.33	0.0005	0.19364917
	FRUCTOSE	2.20000000	11.36	0.0003	0.19364917
	GLUCOSE	2.90000000	14.98	0.0001	0.19364917
	SUCROSE	3.80000000	19.62	0.0001	0.19364917
CONTROL VS SUGARS		-0.96666670	-4.32	0.0124	0.22360674
6-CARBONS VS SUCROSE		1.25000000	5.27	0.0062	0.23717082
FRUCTOSE VS GLUCOSE		0.70000000	2.56	0.0629	0.27386128

means

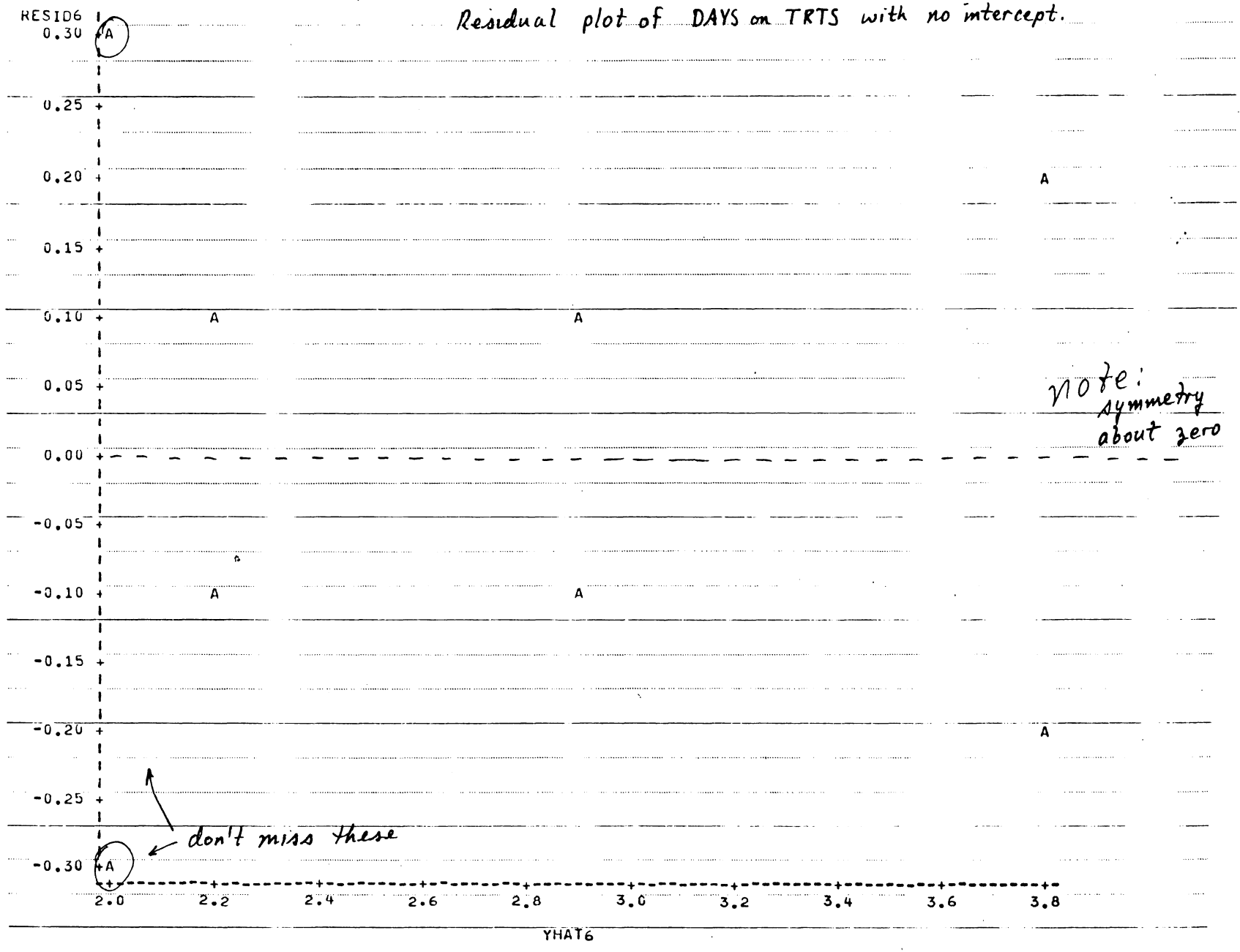
Compare contrast estimates and values with NOTES: pp 16.4, 16.7.

OBSERVATION	OBSERVED VALUE y_i	PREDICTED VALUE $\hat{y}_i$	RESIDUAL $y_i - \hat{y}_i$
1	2.30000000	2.00000000	0.30000000
2	1.70000000	2.00000000	-0.30000000
3	3.60000000	3.80000000	-0.20000000
4	4.00000000	3.80000000	0.20000000
5	3.00000000	2.90000000	0.10000000
6	2.80000000	2.90000000	-0.10000000
7	2.10000000	2.20000000	-0.10000000
8	2.30000000	2.20000000	0.10000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	0.30000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.20000000
DURBIN-WATSON D	2.06666667

LEAFHOPPER DATA

PLOT OF RESID6*YHAT6 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



LEAFHOPPER DATA

OBS	X0	X1	X2	X3	X4
1	1	1	0	0	0
2	1	1	0	0	0
3	1	0	1	0	0
4	1	0	1	0	0
5	1	0	0	1	0
6	1	0	0	1	0
7	1	0	0	0	1
8	1	0	0	0	1

X-matrix for equal means / general means model sequence

LEAFHOPPER DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: DAYS — Regress DAYS on X_1, X_2, X_3, X_4 with intercept.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	3	3.97500000	1.32500000	17.67	0.0090
ERROR	4	0.30000000	0.07500000		
CORRECTED TOTAL	7	4.27500000			

= R(1/0) + R(2/0,1) + R(3/0,1,2) + SSResidual

← does not involve R(0) = 59.405

Compare ANOVA table with NOTES: p.16.6

R-SQUARE	C.V.	STD DEV	DAYS MEAN
0.929825	10.0500	0.27386128	2.72500000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1 alone	1	1.40166667	18.69	0.0124
X2 after X1	1	2.08333333	27.78	0.0062
X3 after X1,X2	1	0.49000000	6.53	0.0629
X4 after X1,X2,X3	0	0.00000000	.	.

cf. NOTES: p.16.6

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	0	0.00000000	.	.
X2	0	0.00000000	.	.
X3	0	0.00000000	.	.
X4	0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	2.20000000 B	11.36	0.0003	0.19364917
X1	-0.20000000 B	-0.73	0.5057	0.27386128
X2	1.60000000 B	5.84	0.0043	0.27386128
X3	0.70000000 B	2.56	0.0629	0.27386128
X4	0.00000000 B	.	.	.

cf. NOTES: p.16.7.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	2.30000000	2.00000000	0.30000000
2	1.70000000	2.00000000	-0.30000000
3	3.60000000	3.80000000	-0.20000000
4	4.00000000	3.80000000	0.20000000
5	3.00000000	2.90000000	0.10000000
6	2.80000000	2.90000000	-0.10000000
7	2.10000000	2.20000000	-0.10000000
8	2.30000000	2.20000000	0.10000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	0.30000000
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.20000000
DURBIN-WATSON D	2.06666667

LYMPHOCYTE DATA

```
2          DATA LYMPHCYT;                    (Unit 17)
3                  INPUT TRTS STIM ATP        Y;
4                  X0=1;
5                  CARDS;                    (See Below)

NOTE: DATA SET WORK.LYMPHCYT HAS 8 OBSERVATIONS AND 6 VARIABLES. 250 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.19 SECONDS AND 116K.

14         TITLE LYMPHOCYTE DATA;
15    (A)  PROC GLM;                         Groups TRTS variable for fitting.
16                 CLASS TRTS;
17                 MODEL Y=TRTS/NOINT SOLUTION P;   Regress Y on TRTS with NO INTercept
18                 OUTPUT OUT=NEW8 RESIDUAL=RESID8 PREDICTED=YHAT8;   (note use of
19         ESTIMATE 'STIMULUS'                                          SOLUTION option).
20              TRTS .5 -.5 .5 -.5/E;
21         ESTIMATE 'ATP PRESENCE'
22              TRTS .5 .5 -.5 -.5/E;     } Estimates contrasts.
23         ESTIMATE 'INTERACTION'
24              TRTS .5 -.5 -.5 .5/E;
25                 MEANS TRTS;                    Calculates TRTS means.

NOTE: DATA SET WORK.NEW8 HAS 8 OBSERVATIONS AND 8 VARIABLES. 191 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.42 SECONDS AND 190K AND PRINTED PAGES 1 TO 7.

    (B)  PROC PLOT;
27                 PLOT RESID8*YHAT8;   Plot residuals of Y regressed on TRTS.

NOTE: THE PROCEDURE PLOT USED 0.18 SECONDS AND 124K AND PRINTED PAGE 8.


30    (C)  PROC GLM;
31                 MODEL Y=STIM ATP ATP*STIM/SOLUTION P;   Regress Y on STIM, ATP and
                                                           ATP*STIM (interaction).
NOTE: THE PROCEDURE GLM USED 0.27 SECONDS AND 162K AND PRINTED PAGES 10 TO 11.
                                                           Note: we don't need residual
NOTE: SAS INSTITUTE INC.                                   plot; residuals are the
      P.O. BOX 10066                                       same as in output (A).
      RALEIGH, N.C. 27605
```

The cards looked like:

TRTS	STIM	ATP	Y	observation
1	0.5	0.5	34	1
1	0.5	0.5	30	2
2	-0.5	0.5	62.5	3
2	-0.5	0.5	67.5	4
3	0.5	-0.5	25.5	5
3	0.5	-0.5	24.5	6
4	-0.5	-0.5	46	7
4	-0.5	-0.5	50	8

LYMPHOCYTE DATA

GENERAL LINEAR MODELS PROCEDURE

Summary of CLASS procedure

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
TRTS	4	1 2 3 4

NUMBER OF OBSERVATIONS IN DATA SET = 8

ESTIMABLE FUNCTIONS FOR STIMULUS

1st contrast vector

EFFECT		COEFFICIENTS
TRTS	1	0.5
	2	-0.5
	3	0.5
	4	-0.5

ESTIMABLE FUNCTIONS FOR ATP PRESENCE

2nd contrast vector

EFFECT		COEFFICIENTS
TRTS	1	0.5
	2	0.5
	3	-0.5
	4	-0.5

ESTIMABLE FUNCTIONS FOR INTERACTION

3rd contrast vector

EFFECT		COEFFICIENTS
TRTS	1	0.5
	2	-0.5
	3	-0.5
	4	0.5

MEANS procedure results

MEANS

TRTS	N	Y
1	2	32.0000000
2	2	65.0000000
3	2	25.0000000
4	2	48.0000000

(A)

LYMPHOCYTE DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y — Regress Y on TRTS (classed) with no intercept.

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	4	16356.00000000	4089.00000000	564.00	0.0001
ERROR	4	29.00000000	7.25000000		
UNCORRECTED TOTAL	8	16385.00000000			

compare with NOTES: p. 17.2.

R-SQUARE	C.V.	STD DEV	Y MEAN
0.998230	6.3355	2.69258240	42.50000000 = $\bar{y}$ (overall)

SOURCE	DF	TYPE I SS	F VALUE	PR > F
TRTS	4	16356.00000000	564.00	0.0001

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
TRTS	4	16356.00000000	564.00	0.0001

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
TRTS	1	32.00000000	16.81	0.0001	1.90394328
	2	65.00000000	34.14	0.0001	1.90394328
	3	25.00000000	13.13	0.0002	1.90394328
	4	48.00000000	25.21	0.0001	1.90394328
STIMULUS		-28.00000000	-14.71	0.0001	1.90394328
ATP PRESENCE		12.00000000	6.30	0.0032	1.90394328
INTERACTION		-5.00000000	-2.63	0.0584	1.90394328

e.g.: compare interaction results with NOTES: pp. 17.8-.9

means

contrast estimates; cf. NOTES: p. 17.6.

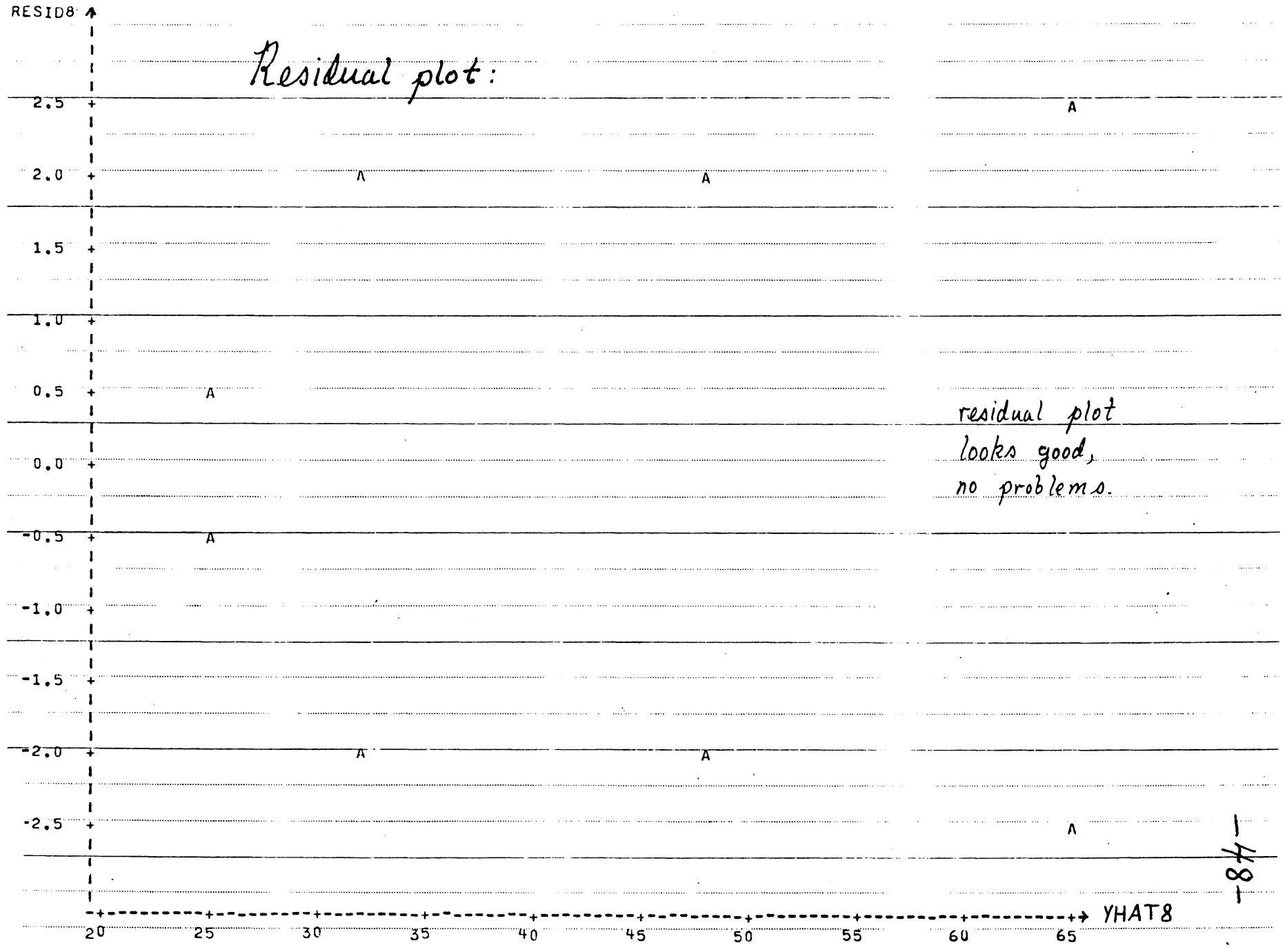
OBSERVATION	OBSERVED VALUE y_i	PREDICTED VALUE $\hat{y}_i$	RESIDUAL $y_i - \hat{y}_i$
1	34.00000000	32.00000000	2.00000000
2	30.00000000	32.00000000	-2.00000000
3	62.50000000	65.00000000	-2.50000000
4	67.50000000	65.00000000	2.50000000
5	25.50000000	25.00000000	0.50000000
6	24.50000000	25.00000000	-0.50000000
7	46.00000000	48.00000000	-2.00000000
8	50.00000000	48.00000000	2.00000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	29.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.25000000
DURBIN-WATSON D	2.22413793

17.1

(B) LYMPHOCYTE A

PLOT OF RESID8*YHAT8 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



residual plot looks good, no problems.

(C)

LYMPHOCYTE DA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

Regress Y on STIM, ATP, and ATP*STIM (interaction)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL Regression	3	1906.00000000	635.33333333	87.63	0.0004
ERROR Residual	4	29.00000000	7.25000000		
CORRECTED TOTAL	7	1935.00000000			

$= R(1/0) + R(2/0,1) + R(3/0,1,2)$

R-SQUARE	C.V.	STD DEV	Y MEAN
0.985013	6.3355	2.69258240	42.50000000

SOURCE	DF	TYPE I SS	F VALUE	PR > F
STIM	1	1568.00000000	216.28	0.0001
ATP	1	288.00000000	39.72	0.0032
STIM*ATP	1	50.00000000	6.90	0.0584

Compare these TYPE I SS results with the contrast estimates in (A).

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
STIM	1	1568.00000000	216.28	0.0001
ATP	1	288.00000000	39.72	0.0032
STIM*ATP	1	50.00000000	6.90	0.0584

Note: since we multiplied ATP by STIM to get the interaction, our X-matrix looks like:

$$X = \begin{bmatrix} 1/2 & 1/2 & 1/4 \\ 1/2 & 1/2 & 1/4 \\ -1/2 & 1/2 & -1/4 \\ -1/2 & 1/2 & -1/4 \\ 1/2 & -1/2 & -1/4 \\ 1/2 & -1/2 & -1/4 \\ -1/2 & -1/2 & 1/4 \\ -1/2 & -1/2 & 1/4 \end{bmatrix}$$

This creates an "error" of a factor of 2 in the STIM*ATP estimate of -10.

Also see NOTES: pp. 17.20a-.21.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	42.50000000	44.64	0.0001	0.95197164
STIM	-28.00000000	-14.71	0.0001	1.90394328
ATP	12.00000000	6.30	0.0032	1.90394328
STIM*ATP	-10.00000000	-2.63	0.0584	3.80788655

cf. NOTES: p. 17.6.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	34.00000000	32.00000000	2.00000000
2	30.00000000	32.00000000	-2.00000000
3	62.50000000	65.00000000	-2.50000000
4	67.50000000	65.00000000	2.50000000
5	25.50000000	25.00000000	0.50000000
6	24.50000000	25.00000000	-0.50000000
7	46.00000000	48.00000000	-2.00000000
8	50.00000000	48.00000000	2.00000000

Same $y_i - \hat{y}_i$ values as in (A).

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	29.00000000
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.25000000
DURBIN-WATSON D	2.22413793

Unit 17 Fat digestibility data.

```
1
2           DATA FAT_DIG;
3           INPUT PERIOD FAT $ LECITIN Y;
4           X1=0;
5           IF PERIOD=1 THEN X1=1;
6           X2=0;
7           IF PERIOD=2 THEN X2=1;
8           X3=1-X1-X2;
9           X4=0;
10          IF FAT='T' AND LECITIN=0 THEN X4=1;
11          X5=0;
12          IF FAT='C' AND LECITIN=0 THEN X5=1;
13          X6=0;
14          IF FAT='T' AND LECITIN=1 THEN X6=1;
15          X7=1-X4-X5-X6;
16          XC4=.5-LECITIN;
17          XC5=X4-X5;
18          XC6=X6-X7;
19          CARDS;
```

creating x-variables. X1-X7 as on p. 17.10 XC4-XC6 are X_4-X_6 on p. 17.13 (contrast variables)

Data cards after CARDS-statement

period fat lecitin y

```
1 T 0 64.6
2 T 0 52.4
3 T 0 53.8
1 C 0 66.0
2 C 0 60.1
3 C 0 64.4
1 T 1 85.0
2 T 1 68.9
3 T 1 77.5
1 C 1 96.0
2 C 1 90.4
3 C 1 98.2
```

NOTE: DATA SET WORK.FAT_DIG HAS 12 OBSERVATIONS AND 14 VARIABLES. 112 OBS/TRK
NOTE: THE DATA STATEMENT USED 0.23 SECONDS AND 116K.

```
32          PROC PRINT;
33    (A)   TITLE FAT DIGESTIBILITY DATA;
34          VAR X1-X7;
```

Print X-matrix for (B) and (D)

NOTE: THE PROCEDURE PRINT USED 0.15 SECONDS AND 116K AND PRINTED PAGE 1.

```
32    (B)   PROC GLM;
33          MODEL Y=X1-X7 /NOINT;
34          ESTIMATE 'W VS WO LECITIN' X4 .5 X5 .5 X6 -.5 X7 -.5;
35    (C)   ESTIMATE 'FAT DIFF WO LEC' X4 1 X5 -1;
36          ESTIMATE 'FAT DIFF W LEC ' X6 1 X7 -1;
```

This model produces same analysis as on p. 17.11 (period fitted first)

contrasts as on p. 17.12

NOTE: THE PROCEDURE GLM USED 0.45 SECONDS AND 166K AND PRINTED PAGES 1 TO 2.

```
37    (D)   PROC GLM;
38          MODEL Y = X1-X7 ;
39          ESTIMATE 'W VS WO LECITIN' X4 .5 X5 .5 X6 -.5 X7 -.5;
40    (C)   ESTIMATE 'FAT DIFF WO LEC' X4 1 X5 -1;
41          ESTIMATE 'FAT DIFF W LEC ' X6 1 X7 -1;
```

Same as in (B) except for common intercept model at bottom of p. 17.13

same contrasts as above (same output)

NOTE: THE PROCEDURE GLM USED 0.45 SECONDS AND 166K AND PRINTED PAGES 3 TO 4.

```
39    (E)   PROC PRINT;
40          VAR X4-X7 X1-X3;
```

Print X-matrix for (F)

NOTE: THE PROCEDURE PRINT USED 0.16 SECONDS AND 116K AND PRINTED PAGE 6.

```
41    (F)   PROC GLM;
42          MODEL Y=X4-X7 X1-X3/P;
```

Model on p. 17.14 period variables fitted last.

NOTE: THE PROCEDURE GLM USED 0.37 SECONDS AND 164K AND PRINTED PAGES 7 TO 8.

(G)
```
PROC PRINT;
VAR X1-X3 XC4-XC6;
```
Print X-matrix for (H)

```
NOTE: THE PROCEDURE PRINT USED 0.22 SECONDS AND 116K AND PRINTED PAGE 1.
```

(H)
```
PROC GLM;
MODEL Y=X1-X3 XC4-XC6/P;
```
Model on pr 17.13 at the top
treatment indicators changed to contrast variables

```
NOTE: THE PROCEDURE GLM USED 0.38 SECONDS AND 164K AND PRINTED PAGES 2 TO 3.
```

(I)
```
46  PROC GLM;
47  CLASSES PERIOD FAT LECITIN;
48  MODEL Y = PERIOD FAT LECITIN FAT*LECITIN/SOLUTION P;
49  ESTIMATE 'W VS WO LECITIN'
50  LECITIN .5 -.5/E;
```
Same results as for (D)

(J)
```
51  ESTIMATE 'FAT DIFF WO LEC'
52  LECITIN 1 FAT 1 -1/E;
53  ESTIMATE 'FAT DIFF W LEC'
54  LECITIN 0 1 FAT 1 -1/E;
55  OUTPUT OUT=NEW5 RESIDUAL=RESID5 PREDICTED=YHAT5;
```
} only this contrast estimable because of redundancies in the variables.

```
NOTE: DATA SET WORK.NEW5 HAS 12 OBSERVATIONS AND 18 VARIABLES. 88 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.54 SECONDS AND 190K AND PRINTED PAGES 11 TO 16.
```

(K)
```
56  PROC PLOT;
57  PLOT RESID5*YHAT5;
```
Residual plot for all models.

```
NOTE: THE PROCEDURE PLOT USED 0.21 SECONDS AND 124K AND PRINTED PAGE 17.
```

FAT DIGESTIBILITY DATA

(F)

X0	OBS	X1	X2	X3	X4	X5	X6	X7
1	1	1	0	0	1	0	0	0
1	2	0	1	0	1	0	0	0
1	3	0	0	1	1	0	0	0
1	4	1	0	0	0	1	0	0
1	5	0	1	0	0	1	0	0
1	6	0	0	1	0	1	0	0
1	7	1	0	0	0	0	1	0
1	8	0	1	0	0	0	1	0
1	9	0	0	1	0	0	1	0
1	10	1	0	0	0	0	0	1
1	11	0	1	0	0	0	0	1
1	12	0	0	1	0	0	0	1

X matrix for model in (B)

Same matrix with the X0 column in front is the X matrix for model in (D)

(B) See p. 17.11

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

Model $y = x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7$ / noint

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	6	66867.48166667	11144.58027778	940.36	0.0001
ERROR	6	71.10833333	11.85138889		
UNCORRECTED TOTAL	12	66938.59000000	= Total SS		= residual SS

R-SQUARE	C.V.	STD DEV	Y MEAN
0.998938	4.7089	3.44258462	73.10833333

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	24273.64000000	2048.17	0.0001
X2	1	18468.81000000	1558.37	0.0001
X3	1	21594.30250000	1822.09	0.0001
X4	1	1046.52250000	88.30	0.0001
X5	1	1012.50000000	85.43	0.0001
X6	1	471.70666667	39.80	0.0007
X7	0	0.00000000	.	.

Periods SS = R(1,2,3) = 64336.75250000

Treatment SS = R(4,5,6 | 1,2,3) = 2530.72916667

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	0	0.00000000	.	.
X2	0	0.00000000	.	.
X3	0	0.00000000	.	.
X4	0	0.00000000	.	.
X5	0	0.00000000	.	.
X6	0	0.00000000	.	.
X7	0	0.00000000	.	.

Same as table p 17.11

PARAMETER	ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1 period means	99.65833333	B	40.94	0.0001	2.43427493
X2 + $(\bar{y}_4 - \bar{y})$	89.70833333	B	36.85	0.0001	2.43427493
X3	95.23333333	B	39.12	0.0001	2.43427493
X4 $\bar{y}_1 - \bar{y}_4$ =	-37.93333333	B	-13.50	0.0001	2.81085857
X5 $\bar{y}_2 - \bar{y}_4$ =	-31.36666667	B	-11.16	0.0001	2.81085857
X6 $\bar{y}_3 - \bar{y}_4$ =	-17.73333333	B	-6.31	0.0007	2.81085857
X7 $\bar{y}_4 - \bar{y}_4$ =	0.00000000	B	.	.	.

21.758333

77.9 + 94.866666 − 78.108333

67.96 + 21.758333

73.475 + 21.758333

last contrast on p. 17.12

$b_4 - b_5 = -6.56666666$ estimating $\bar{y}_1 - \bar{y}_2$ second contrast on p 17.12

$\frac{1}{2}b_4 + \frac{1}{2}b_5 - \frac{1}{2}b_6 - \frac{1}{2}b_7 = -25.78333333$ first contrast on p 17.12

(D)

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

model $Y = X_1\ X_2\ X_3\ X_4\ X_5\ X_6\ X_7$

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	5	2729.54083333	545.90816667	46.06	0.0001
ERROR	6	71.10833333	11.85138889		
CORRECTED TOTAL	11	2800.64916667	= 66938.59 − 12 $\bar{Y}^2$ = 66938.59 − 64137.9408 3333		

R-SQUARE	C.V.	STD DEV	Y MEAN
0.974610	4.7089	3.44258462	73.10833333

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X1	1	137.76041667	11.62	0.0143
X2	1	61.05125000	5.15	0.0637
X3	0	0.00000000	.	.
X4	1	1046.52250000	88.30	0.0001
X5	1	1012.50000000	85.43	0.0001
X6	1	471.70666667	39.80	0.0007
X7	0	0.00000000	.	.

Period SS = R(1,2|0) = 198.81166667
= 64336.7525 − 64137.9408 3333

Treatment SS = R(4,5,6|1,2,3)
= 2530.72916667

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	0	0.00000000	.	.
X2	0	0.00000000	.	.
X3	0	0.00000000	.	.
X4	0	0.00000000	.	.
X5	0	0.00000000	.	.
X6	0	0.00000000	.	.
X7	0	0.00000000	.	.

as on (B)

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT 3. per mean ($+\bar{y}_4 - \bar{y}$)	95.23333333 B	39.12	0.0001	2.43427493
X1 1.−3. per. mean	4.42500000 B	1.82	0.1190	2.43427493
X2 2.−3. per. mean	−5.52500000 B	−2.27	0.0637	2.43427493
X3 3.−3. per. mean	0.00000000 B	.	.	.
X4 $\bar{y}_1 - \bar{y}_4$	−37.93333333 B	−13.50	0.0001	2.81085857
X5 $\bar{y}_2 - \bar{y}_4$	−31.36666667 B	−11.16	0.0001	2.81085857
X6 $\bar{y}_3 - \bar{y}_4$	−17.73333333 B	−6.31	0.0007	2.81085857
X7 $\bar{y}_4 - \bar{y}_4$	0.00000000 B	.	.	.

5

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE USED FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE FIRST ESTIMATOR AND THE T VALUE TESTS HO: (BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

PARAMETER	ESTIMATE	T FOR HO: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
W VS WO LECITIN	-25.73333333	-12.97	0.0001	1.98377716
FAT DIFF WO LEC	-6.56666667	-2.34	0.0581	2.81085857
FAT DIFF W LEC	-17.73333333	-6.31	0.0007	2.81085857

Compare with table on p. 17.12

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

P-option for all models (B) (D) (F) (H) and (I) all produces this part

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS HO: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	64.60000000	61.72500000	2.87500000
2	52.40000000	51.77500000	0.62500000
3	53.80000000	57.30000000	-3.50000000
4	66.00000000	68.29166667	-2.29166667
5	60.10000000	58.34166667	1.75833333
6	64.40000000	63.86666667	0.53333333
7	85.00000000	81.92500000	3.07500000
8	68.90000000	71.97500000	-3.07500000
9	77.50000000	77.50000000	0.00000000
10	96.00000000	99.65833333	-3.65833333
11	90.40000000	89.70833333	0.69166667
12	98.20000000	95.23333333	2.96666667

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	71.10833333
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.05281456
DURBIN-WATSON D	1.86561877

OBS	X0	X4	X5	X6	X7	X1	X2	X3
1	1	1	0	0	0	1	0	0
2	1	1	0	0	0	0	1	0
3	1	1	0	0	0	0	0	1
4	1	0	1	0	0	1	0	0
5	1	0	1	0	0	0	1	0
6	1	0	1	0	0	0	0	1
7	1	0	0	1	0	1	0	0
8	1	0	0	1	0	0	1	0
9	1	0	0	1	0	0	0	1
10	1	0	0	0	1	1	0	0
11	1	0	0	0	1	0	1	0
12	1	0	0	0	1	0	0	1

X matrix for model in (F)

(F) See p. 17.14

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

Model $y = X_4 X_5 X_6 X_7 X_1 X_2 X_3$

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	5	2729.54083333	545.90816667	46.06	0.0001
ERROR	6	71.10833333	11.85138889		
CORRECTED TOTAL	11	2800.64916667			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.974610	4.7089	3.44258462	73.10833333

Sequential SS

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X4	1	1046.52250000	88.30	0.0001
X5	1	1012.50000000	85.43	0.0001
X6	1	471.70666667	39.80	0.0007
X7	0	0.00000000	.	.
X1	1	137.75041667	11.62	0.0143
X2	1	61.05125000	5.15	0.0637
X3	0	0.00000000	.	.

Treatment SS = R(4,5,6 | 0)
= 2530.72916667

Period SS = R(1,2 | 4,5,6)
= 198.81166667

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X4	0	0.00000000	.	.
X5	0	0.00000000	.	.
X6	0	0.00000000	.	.
X7	0	0.00000000	.	.
X1	0	0.00000000	.	.
X2	0	0.00000000	.	.
X3	0	0.00000000	.	.

Compare with (C) Note that the sequential SS is not depend on the order of the treatment and period indicator variables as mentioned on p. 17.15

PARAMETER	ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	95.23333333	B $= \bar{y}_{P2} + \bar{y}_{T1} - \bar{y}$	39.12	0.0001	2.43427493
X4	-37.93333333	B $= \bar{y}_{T1} - \bar{y}_{T4}$	-13.50	0.0001	2.81085857
X5	-31.36666667	B $= \bar{y}_{T2} - \bar{y}_{T4}$	-11.16	0.0001	2.81085857
X6	-17.73333333	B $= \bar{y}_{T3} - \bar{y}_{T4}$	-6.31	0.0007	2.81085857
X7	0.00000000	B	.	.	.
X1	4.42500000	B $= \bar{y}_{P1} - \bar{y}_{P3}$	1.82	0.1190	2.43427493
X2	-5.52500000	B $= \bar{y}_{P2} - \bar{y}_{P4}$	-2.27	0.0637	2.43427493
X3	0.00000000	B	.	.	.

(G)

S T A T I S T I C A L A N A L Y S I S S Y S T E M

OBS	Xo	X_1	X_2	X_3	XC4	XC5	XC6
1	1	1	0	0	0.5	1	0
2	1	0	1	0	0.5	1	0
3	1	0	0	1	0.5	1	0
4	1	1	0	0	0.5	-1	0
5	1	0	1	0	0.5	-1	0
6	1	0	0	1	0.5	-1	0
7	1	1	0	0	-0.5	0	1
8	1	0	1	0	-0.5	0	1
9	1	0	0	1	-0.5	0	1
10	1	1	0	0	-0.5	0	-1
11	1	0	1	0	-0.5	0	-1
12	1	0	0	1	-0.5	0	-1

X matrix for model in (H) on next page

(H)

DEPENDENT VARIABLE:

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	5	2729.[illegible]3333[illegible]	[illegible]	[illegible]	0.0001
ERROR	6	71.10833333	11.85138889		
CORRECTED TOTAL	11	2800.64916667			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.974610	[illegible]	3.44258462	73.10833333

SOURCE	DF	TYPE I SS	F VALUE	PR > F	
X1	1	137.76041667	11.62	0.0143	} same as (D)
X2	1	61.05125000	5.15	0.0637	
X3	0	0.00000000	.	.	
XC4	1	1994.34033333	168.28	0.0001	} same as for X4-X6 on (D)
XC5	1	64.68166667	5.46	0.0581	
XC6	1	471.70666667	39.80	0.0007	

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X1	0	0.00000000	.	.
X2	0	0.00000000	.	.
X3	0	0.00000000	.	.
XC4	1	1994.34033333	168.28	0.0001
XC5	1	64.68166667	5.46	0.0581
XC6	1	471.70666667	39.80	0.0007

note: no redundancies here (only 3 variables)

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	$\bar{y}_{P3}$ = 73.47500000 C	42.69	0.0001	1.72129231
X1	$\bar{y}_{P1} - \bar{y}_{P3}$ = 4.42500000 B	1.[illegible]	0.1180	2.45487493
X2	$\bar{y}_{P2} - \bar{y}_{P1}$ = -5.57500000 B	-2.27	0.[illegible]	2.45487493
X3	0.00000000 B	.	.	.
XC4	-25.78333333	-12.97	0.0001	1.98757716
XC5	-3.28333333	-2.34	0.0581	1.40542929
XC6	-8.86666667	-6.31	0.0007	1.40542929

and these produce table on p 17.12

multiply by 2

151

FAT DIGESTIBILITY DATA

(I)

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
PERIOD	3	1 2 3
FAT	2	C T
LECITIN	2	0 1

Produced by the classes statement

NUMBER OF OBSERVATIONS IN DATA SET = 12

(J)

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR W VS WO LECITIN

EFFECT		COEFFICIENTS
INTERCEPT		0
PERIOD	1	0
	2	0
	3	0
FAT	C	0
	T	0
LECITIN	0	0.5
	1	-0.5
FAT*LECITIN	C 0	0.25
	C 1	-0.25
	T 0	0.25
	T 1	-0.25

1st estimate statement

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR FAT DIFF WO LEC

EFFECT		NON-EST COEFFICIENTS
INTERCEPT		0
PERIOD	1	0
	2	0
	3	0
FAT	C	1
	T	-1
LECITIN	0	1
	1	0
FAT*LECITIN	C 0	1
	C 1	0.5
	T 0	0
	T 1	-0.5

2nd estimate statement

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR FAT DIFF W LEC

EFFECT		NON-EST COEFFICIENTS
INTERCEPT		0
PERIOD	1	0
	2	0
	3	0
FAT	C	1
	T	-1
LECITIN	0	0
	1	1
FAT*LECITIN	C 0	0.5
	C 1	1
	T 0	-0.5
	T 1	0

3rd estimate statement

(I)

FAT DIGESTIBILITY DATA

GENERAL LINEAR MODELS PROCEDURE

Model y = period fat lecitin fat * lecitin

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	5	2729.54083333	545.90816667	46.06	0.0001
ERROR	6	71.10833333	11.85138889		
CORRECTED TOTAL	11	2800.64916667			

R-SQUARE	C.V.	STD DEV	Y MEAN
0.974610	4.7089	3.44258462	73.10833333

See SS

SOURCE	DF	TYPE I SS	F VALUE	PR > F
PERIOD	2	198.81166667	8.39	0.0183
FAT	1	442.86750000	37.37	0.0009
LECITIN	1	1994.34083333	168.28	0.0001
FAT*LECITIN	1	93.52083333	7.89	0.0308

= R(1,2,3,4,5)

Treatment SS = R(4,5 | 1,2,3)
= 2530.729 16666

Part SS

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
PERIOD	2	198.81166667	8.39	0.0183
FAT	1	442.86750000	37.37	0.0009
LECITIN	1	1994.34083333	168.28	0.0001
FAT*LECITIN	1	93.52083333	7.89	0.0308

Note partial and sequential SS are the same

PARAMETER		ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		77.50000000	B	31.84	0.0001	2.43427493
PERIOD	1	4.42500000	B	1.82	0.1190	2.43427493
	2	-5.52500000	B	-2.27	0.0637	2.43427493
	3	0.00000000	B	.	.	.
FAT	C	17.73333333	B	6.31	0.0007	2.81085857
	T	0.00000000	B	.	.	.
LECITIN	0	-20.20000000	B	-7.19	0.0004	2.81085857
	1	0.00000000	B	.	.	.
FAT*LECITIN	C 0	-11.16666667	B	-2.81	0.0308	3.97515432
	C 1	0.00000000	B	.	.	.
	T 0	0.00000000	B	.	.	.
	T 1	0.00000000	B	.	.	.

(J)

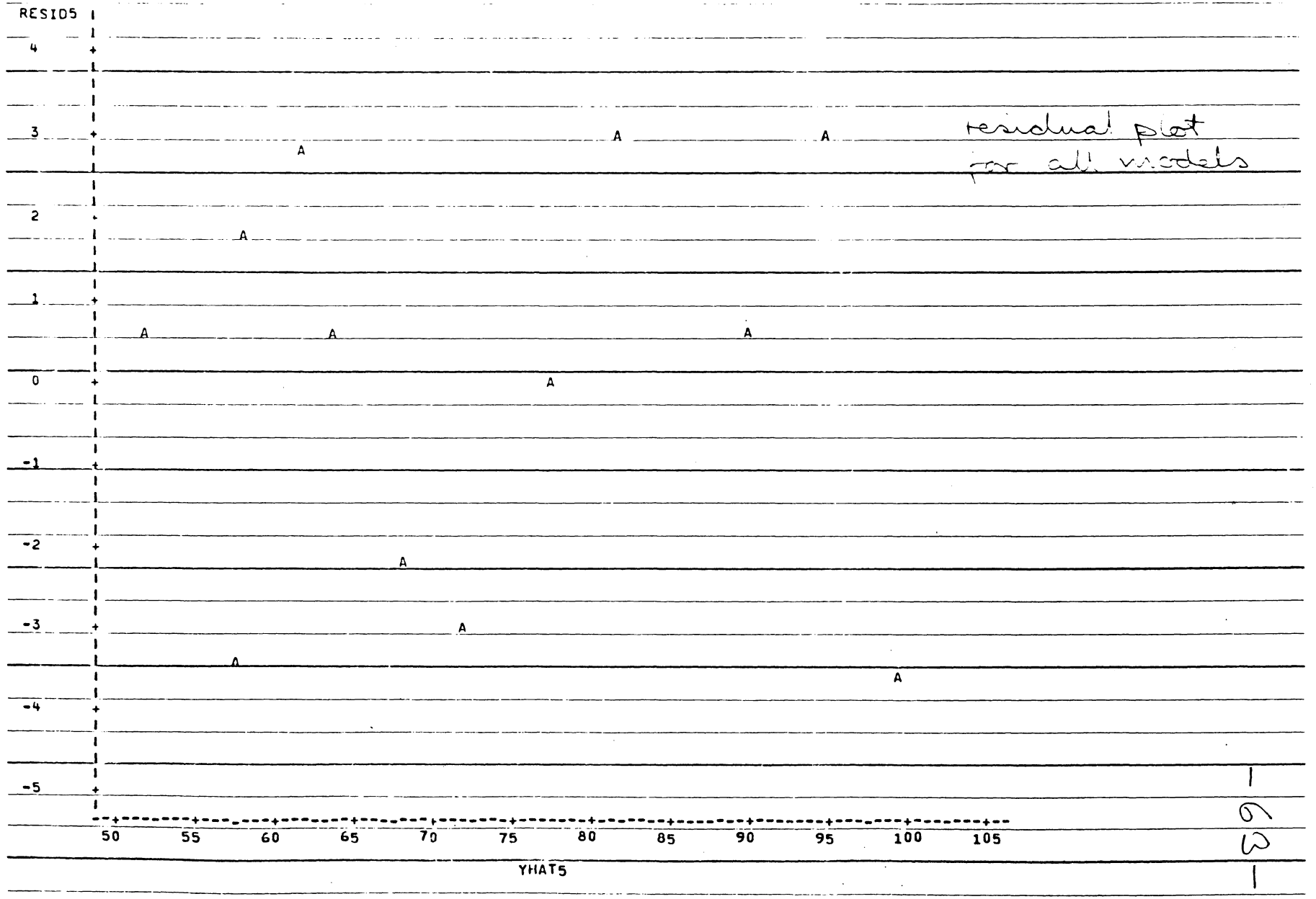
PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
W VS WO LECITIN	-12.89166667	-12.97	0.0001	0.99378858
FAT DIFF WO LEC	NON-EST			
FAT DIFF W LEC	NON-EST			

1
0
2
1

(K)

FAT DIGESTIBILITY DATA

PLOT OF RESID5*YHAT5 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



residual plot for all models

16

```
1                                  S T A T I S T I C A L   A N A L Y S I S   S

NOTE: THE JOB VMPJ0001 HAS BEEN RUN UNDER RELEASE 79.2B OF SAS AT CORNELL UNIVERSITY
                    PROTEIN NUTRITION DATA                          DATA :  PROTEIN Y
1                                                                           H 179
2           DATA PROTEIN;                                                   H 160
3           INPUT PROTEIN $ Y;                                              H 136
4           X0=1;                                                           H 227
5           X1=0;                                                           H 217
6           X2=0;                                                           H 168
7           IF PROTEIN='H' THEN X1=1;                                       H 108
8           IF PROTEIN='L' THEN X2=1;                                       H 124
9           X3=1-X1-X2;                                                     H 143
10          X4=X1-.5*(X2+X3);                                               H 140
11          X5=X2-X3;                                                       L 309
12          X6=13/18*X1-5/18*(X2+X3);                                       L 229
13          X7=7/13*X2-6/13*X3;                                             L 181
14          CARDS;                                                          L 141
                                                                            L 260
NOTE: DATA SET WORK.PROTEIN HAS 36 OBSERVATIONS AND 10 VARIABLES. 155 OBS   L 203
NOTE: THE DATA STATEMENT USED 0.22 SECONDS AND 116K.                        L 148
                                                                            L 169
51          PROC PRINT;                                                     L 213
52          TITLE PROTEIN NUTRITION DATA;                                   L 257
53    (A)   VAR X1-X3;                                                     L 244
54          VAR X0 X4 X5;                                                   L 271
55          VAR X0 X6 X7;                                                   S 243
                                                                            S 230
NOTE: THE PROCEDURE PRINT USED 0.23 SECONDS AND 116K AND PRINTED PAGE 1.    S 248
                                                                            S 327
56    (B)   PROC GLM;                                                      S 329
57          MODEL Y=X4 X5/P*;                                               S 250
                                                                            S 193
NOTE: THE PROCEDURE GLM USED 0.36 SECONDS AND 162K AND PRINTED PAGES 2 TO   S 271
                                                                            S 316
58    (C)   PROC GLM;                                                      S 267
59          MODEL Y=X6 X7/P*;                                               S 199
                                                                            S 177
NOTE: THE PROCEDURE GLM USED 0.33 SECONDS AND 162K AND PRINTED PAGES 4 TO   S 158
                                                                            S 248
60          PROC GLM;
61    (D)   CLASS PROTEIN;
62          MODEL Y = PROTEIN/NOINT SOLUTION P*;
63          ESTIMATE 'HORSEBEAN VS OILMEAL'
64          PROTEIN 1. -.5 -.5/E;
65          ESTIMATE 'LINSEED VS SOYBEAN'
66          PROTEIN 0. 1. -1./E;
67          ESTIMATE 'ORTHO H-B VS OILMEAL'
68          PROTEIN 1. -.461538 -.538462/E;
69          OUTPUT OUT=NEW RESIDUAL=RESID PREDICTED=YHAT;

NOTE: DATA SET WORK.NEW HAS 36 OBSERVATIONS AND 12 VARIABLES. 130 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 188K AND PRINTED PAGES 6 TO 11.

70    (F)   PROC PLOT;
71          PLOT RESID*YHAT;

NOTE: THE PROCEDURE PLOT USED 0.20 SECONDS AND 124K AND PRINTED PAGE 12.
NOTE: SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

constructing treatment indicators

constructing contrast variables

Printing X-matrices for the 3 models: (B), (C) and (D)

model with natural contrasts.

model with ORTHO contrast.

Means model

natural contrasts as in (B)

ORTHO contrasts as in (C)

residual plot for all models

* ALL P-options produce (E)

PROTEIN NUTRITION DATA

(A)

OBS	X1	X2	X3	X0	X4	X5	X0	X6	X7
1	1	0	0	1	1.0	0	1	0.72222	0.00000
2	1	0	0	1	1.0	0	1	0.72222	0.00000
3	1	0	0	1	1.0	0	1	0.72222	0.00000
4	1	0	0	1	1.0	0	1	0.72222	0.00000
5	1	0	0	1	1.0	0	1	0.72222	0.00000
6	1	0	0	1	1.0	0	1	0.72222	0.00000
7	1	0	0	1	1.0	0	1	0.72222	0.00000
8	1	0	0	1	1.0	0	1	0.72222	0.00000
9	1	0	0	1	1.0	0	1	0.72222	0.00000
10	1	0	0	1	1.0	0	1	0.72222	0.00000
11	0	1	0	1	-0.5	1	1	-0.27778	0.53846
12	0	1	0	1	-0.5	1	1	-0.27778	0.53846
13	0	1	0	1	-0.5	1	1	-0.27778	0.53846
14	0	1	0	1	-0.5	1	1	-0.27778	0.53846
15	0	1	0	1	-0.5	1	1	-0.27778	0.53846
16	0	1	0	1	-0.5	1	1	-0.27778	0.53846
17	0	1	0	1	-0.5	1	1	-0.27778	0.53846
18	0	1	0	1	-0.5	1	1	-0.27778	0.53846
19	0	1	0	1	-0.5	1	1	-0.27778	0.53846
20	0	1	0	1	-0.5	1	1	-0.27778	0.53846
21	0	1	0	1	-0.5	1	1	-0.27778	0.53846
22	0	1	0	1	-0.5	1	1	-0.27778	0.53846
23	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
24	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
25	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
26	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
27	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
28	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
29	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
30	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
31	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
32	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
33	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
34	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
35	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154
36	0	0	1	1	-0.5	-1	1	-0.27778	-0.46154

X-matrix for model in (D)

X-matrix for model in (B)

X-matrix for model in (C)

(B)

PROTEIN NUTRITION DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	44317.18571429	22158.59285714	9.07	0.0007
ERROR	33	80659.56428571	2444.22922078		
CORRECTED TOTAL	35	124976.75000000			

Natural contrast variables model

R-SQUARE	C.V.	STD DEV	Y MEAN
0.354603	23.1656	49.43914664	213.41666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X4	1	39212.49615385	16.04	0.0003
X5	1	5104.68956044	2.09	0.1578

see ANOVA table p.18.6

different seq x part SS

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X4	1	38007.50223789	15.55	0.0004
X5	1	5104.68956044	2.09	0.1578

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	208.60238095	25.08	0.0001	8.31796244
X4	-48.40238095 × 1.5 = -72.60	-3.94	0.0004	12.27447785
X5	-14.05357143 × 2 = -28.11	-1.45	0.1578	9.72462814

compare with table p.18.3

(C)

PROTEIN NUTRITION DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	2	44317.18571429	22158.59285714	9.07	0.0007
ERROR	33	80659.56428571	2444.22922078		
CORRECTED TOTAL	35	124976.75000000			

ORTHO contrast variables model

R-SQUARE	C.V.	STD DEV	Y MEAN
0.354603	23.1656	49.43914664	213.41666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F
X6	1	39212.49615385	16.04	0.0003
X7	1	5104.68956044	2.09	0.1578

Type I & IV SS are the same.

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
X6	1	39212.49615385	16.04	0.0003
X7	1	5104.68956044	2.09	0.1578

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	213.41666667	25.90	0.0001	8.23985777
X6	-73.68461538	-4.01	0.0003	18.39651430
X7	-28.10714286	-1.45	0.1578	19.44925628

See table p. 18.3 and on top of p. 18.6

PROTEIN NUTRITION DATA

GENERAL LINEAR MODELS PROCEDURE

(D)

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
PROTEIN	3	H L S

NUMBER OF OBSERVATIONS IN DATA SET = 36

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR HORSEBEAN VS OILMEAL

EFFECT		COEFFICIENTS
PROTEIN	H	1
	L	-0.5
	S	-0.5

natural contrast

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR LINSEED VS SOYBEAN

EFFECT		COEFFICIENTS
PROTEIN	H	0
	L	1
	S	-1

natural and ORTHO contrast

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR ORTHO H-B VS OILMEAL

EFFECT		COEFFICIENTS
PROTEIN	H	1
	L	-0.461538
	S	-0.538462

ORTHO contrast

(D)

PROTEIN NUTRITION DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F
MODEL	3	1683997.43571429	561332.47857143	229.66	0.0001
ERROR	33	80659.56428571	2444.22922078		
UNCORRECTED TOTAL	36	1764657.00000000			

Means model

R-SQUARE	C.V.	STD DEV	Y MEAN
0.954292	23.1656	49.43914664	213.41666667

SOURCE	DF	TYPE I SS	F VALUE	PR > F
PROTEIN	3	1683997.43571429	229.66	0.0001

SOURCE	DF	TYPE IV SS	F VALUE	PR > F
PROTEIN	3	1683997.43571429	229.66	0.0001

PARAMETER			ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
PROTEIN	H	$\bar{y}_1$ =	160.20000000	10.25	0.0001	15.63403090
	L	$\bar{y}_2$ =	218.75000000	15.33	0.0001	14.27185231
	S	$\bar{y}_3$ =	246.85714286	18.68	0.0001	13.21316773
HORSEBEAN VS OILMEAL			-72.60357143	-3.94	0.0004	18.41171677
LINSEED VS SOYBEAN			-28.10714286	-1.45	0.1578	19.44925628
ORTHO H-B VS OILMEAL			-73.68462836	-4.01	0.0003	18.39651430

→ compare with estimates in (B) & (C) and table p.18.3

If model $y = x_0 - x_3$ was run you get ANOVA of p 18.6 and estimates :

b_0	$\bar{y}_3$ =	246.85714286
b_1	$\bar{y}_1 - \bar{y}_3$ =	-86.65714286
b_2	$\bar{y}_2 - \bar{y}_3$ =	-28.10714286
b_3	$\bar{y}_3 - \bar{y}_3$ =	0.0

as on p. 18.7 .

(E)

For all models:

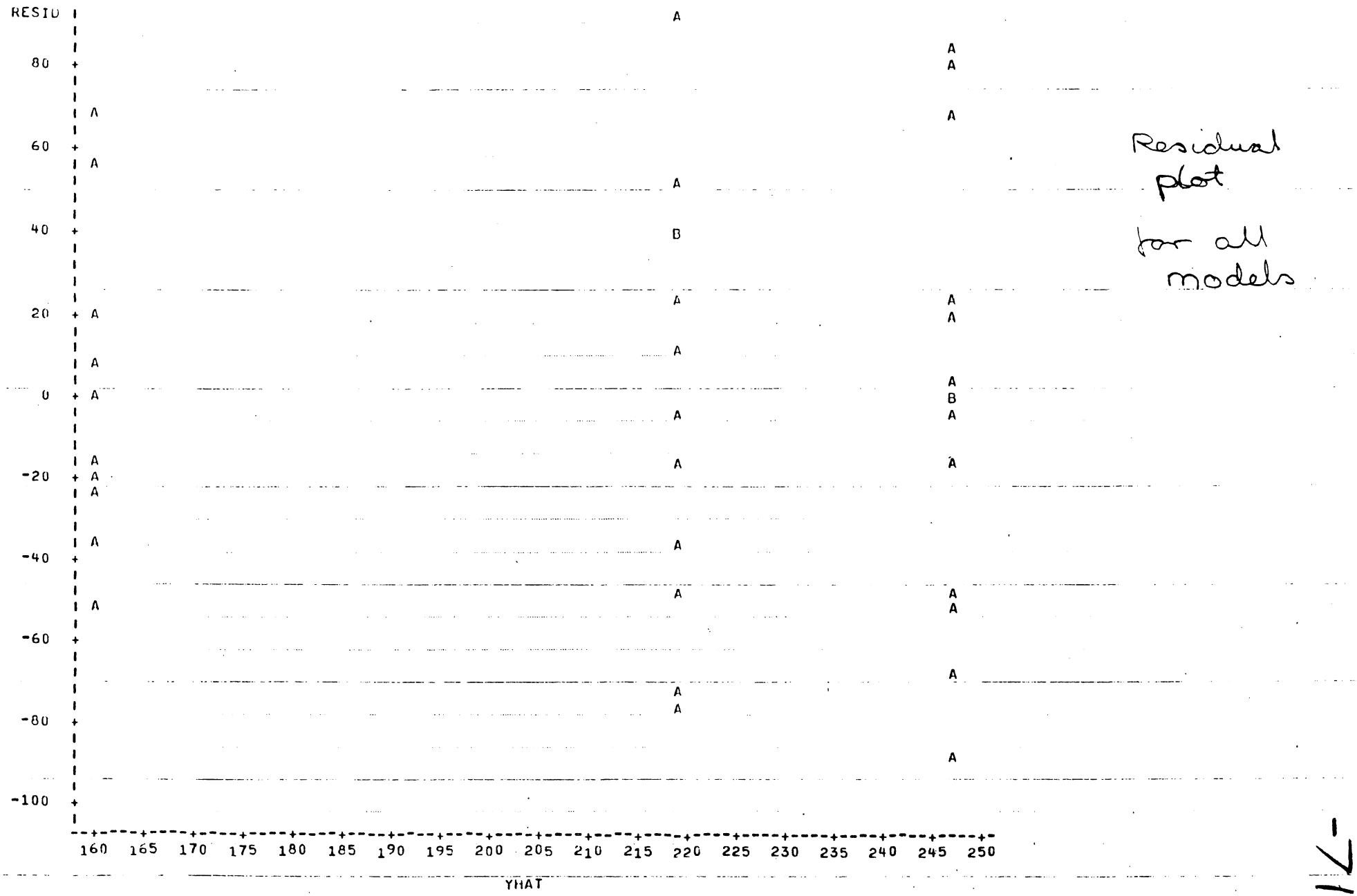
OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	179.00000000	160.20000000	18.80000000
2	160.00000000	160.20000000	-0.20000000
3	136.00000000	160.20000000	-24.20000000
4	227.00000000	160.20000000	66.80000000
5	217.00000000	160.20000000	56.80000000
6	168.00000000	160.20000000	7.80000000
7	108.00000000	160.20000000	-52.20000000
8	124.00000000	160.20000000	-36.20000000
9	143.00000000	160.20000000	-17.20000000
10	140.00000000	160.20000000	-20.20000000
11	309.00000000	218.75000000	90.25000000
12	229.00000000	218.75000000	10.25000000
13	181.00000000	218.75000000	-37.75000000
14	141.00000000	218.75000000	-77.75000000
15	260.00000000	218.75000000	41.25000000
16	203.00000000	218.75000000	-15.75000000
17	148.00000000	218.75000000	-70.75000000
18	169.00000000	218.75000000	-49.75000000
19	213.00000000	218.75000000	-5.75000000
20	257.00000000	218.75000000	38.25000000
21	244.00000000	218.75000000	25.25000000
22	271.00000000	218.75000000	52.25000000
23	243.00000000	246.85714286	-3.85714286
24	230.00000000	246.85714286	-16.85714286
25	248.00000000	246.85714286	1.14285714
26	327.00000000	246.85714286	80.14285714
27	329.00000000	246.85714286	82.14285714
28	250.00000000	246.85714286	3.14285714
29	193.00000000	246.85714286	-53.85714286
30	271.00000000	246.85714286	24.14285714
31	316.00000000	246.85714286	69.14285714
32	267.00000000	246.85714286	20.14285714
33	199.00000000	246.85714286	-47.85714286
34	177.00000000	246.85714286	-69.85714286
35	158.00000000	246.85714286	-88.85714286
36	248.00000000	246.85714286	1.14285714

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	80659.56428571
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	0.33111491
DURBIN-WATSON D	1.33337212

F

PROTEIN NUTRITION DATA

PLOT OF RESID*YHAT LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Residual plot for all models

1 STATISTICAL ANALYSIS

NOTE: THE JOB VMPJ0002 HAS BEEN RUN UNDER RELEASE 79.2B OF SAS AT CORNELL UNIVERSIT

SWAMP PH DATA

```
1          DATA SWAMP_PH;
2          INPUT STRM_LOC $ SWMP_COM $ Y;
3          X0=1;
4          X1=(STRM_LOC='N');   } row indicators
5          X2=1-X1;             }
6          X3=(SWMP_COM='N');   }
7          X4=(SWMP_COM='M');   } column ind's.
8          X5=1-X3-X4;          }
9          X6=X1*X3;            }
10         X7=X1*X4;            }
11         X8=X1*X5;            } interaction indicators
12         X9=X2*X3;            }
13         X10=X2*X4;           }
14         X11=X2*X5;           }
15         CARDS;
```

stream loc. swamp com. Y

```
N N 6.6
N N 7.2
N N 7.2
N N 7.0
N N 6.8
N N 6.4
N N 7.0
N N 6.4
N M 6.8
N M 7.0
N M 6.2
N M 6.2
N M 6.4
N M 6.2
N S 6.4
```

```
NOTE: DATA SET WORK.SWAMP_PH HAS 35 OBSERVATIONS AND 15 VARIABLES. 105 OB
NOTE: THE DATA STATEMENT USED 0.25 SECONDS AND 116K.
```

```
N S 5.2
N S 6.2
N S 6.4
```

```
51         PROC GLM;
52         TITLE SWAMP PH DATA;
53    (A)  CLASSES STRM_LOC SWMP_COM;
54         MODEL Y = STRM_LOC SWMP_COM STRM_LOC*SWMP_COM/SOLUTION P;
55         OUTPUT OUT=NEWA RESIDUAL=RESIDA PREDICTED=YHATA;
56    (B)  P-option prints observed predicted & residual values.
```

```
N S 6.4
A N 6.8
A N 7.0
A M 6.2
A M 5.6
A M 7.2
```

```
NOTE: DATA SET WORK.NEWA HAS 35 OBSERVATIONS AND 17 VARIABLES. 93 OBS/TRK
NOTE: THE PROCEDURE GLM USED 0.45 SECONDS AND 190K AND PRINTED PAGES 1 TO
```

```
A M 7.2
A M 7.2
```

```
57    (C)  PROC PLOT;
58         PLOT RESIDA*YHATA;
```

Residual plot for models (A), (D), (F) & (H).

```
A M 6.2
A M 7.2
A M 7.2
A S 7.2
```

```
NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 126K AND PRINTED PAGE 4.
```

```
A S 6.8
A S 7.0
```

```
59         PROC GLM;
60    (D)  CLASSES STRM_LOC SWMP_COM;
61         MODEL Y = SWMP_COM STRM_LOC STRM_LOC*SWMP_COM/SOLUTION P;
62         ESTIMATE 'ROW DIFFERENCES'
63         STRM_LOC 1. -1./E;
64         ESTIMATE 'COLUMN DIFFERENCE 1'
65         SWMP_COM 0. 1. -1./E;
66         ESTIMATE 'COLUMN DIFFERENCE 2'
67         SWMP_COM 1. -.5 -.5/E;
```

see (C)

```
A S 6.8
A S 7.0
A S 6.6
```

```
NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 165K AND PRINTED PAGES 5 TO 9.
```

```
68    (E)  PROC PRINT;
69         VAR X0-X11;
```

Print X-matrix of model in (F)

```
NOTE: THE PROCEDURE PRINT USED 0.25 SECONDS AND 116K AND PRINTED PAGE 10.
```

```
70    (F)  PROC GLM;
71         MODEL Y=X1-X11;
```

equal means | row ind's | column ind's | general means
same as model in (A)

```
NOTE: THE PROCEDURE GLM USED 0.34 SECONDS AND 164K AND PRINTED PAGE 11.

72         PROC PRINT;
```

```
71    (G) VAR X0 X3-X5 X1 X2 X6-X11;
```
Print X-matrix for model in (H)

```
NOTE: THE PROCEDURE PRINT USED 0.25 SECONDS AND 116K AND PRINTED PAGE 12.

72    (H) PROC GLM;
73        MODEL Y=X3-X5 X1 X2 X6-X11;
```
equal means / column ind's / row ind's / same as model in (D) general means

```
NOTE: THE PROCEDURE GLM USED 0.42 SECONDS AND 166K AND PRINTED PAGE 13.

74    (I) PROC PRINT;
75        VAR X0-X5;
```
Print X-matrix for model in (J)

```
NOTE: THE PROCEDURE PRINT USED 0.21 SECONDS AND 116K AND PRINTED PAGE 14.

76    (J) PROC GLM;
77        MODEL Y=X1-X5/P;
```
equal means / row ind's / column ind's / a reduced model of that in (F)

```
NOTE: THE PROCEDURE GLM USED 0.29 SECONDS AND 162K AND PRINTED PAGE 15.

78    (K) PROC PRINT;
79        VAR X0 X3-X5 X1 X2;
```
Print X-matrix for model in (L)

```
NOTE: THE PROCEDURE PRINT USED 0.21 SECONDS AND 116K AND PRINTED PAGE 16.

80    (L) PROC GLM;
81        MODEL Y=X3-X5 X1 X2;
```
equal means / column ind's / row ind's / a reduced model of that in (H)

```
NOTE: THE PROCEDURE GLM USED 0.28 SECONDS AND 162K AND PRINTED PAGE 17.

NOTE: SAS INSTITUTE INC.
      P.O. BOX 10066
      RALEIGH, N.C. 27605
```

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
STRM_LOC	2	A N
SWMP_COM	3	M N S

NUMBER OF OBSERVATIONS IN DATA SET = 35

(A)

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	2.42538095	0.48507619	2.50	0.0536	0.300852	6.6167
ERROR	29	5.63633333	0.19435632		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429 = Total SS − R(M)			0.44085862		6.66285714

ANOVA(I) at p. 18.16

R(M) = 35 × 7² = 1543.79

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
STRM_LOC	1	R(L|M) = 0.77487218	3.99	0.0553	1	1.00971645	5.20	0.0302
SWMP_COM	2	R(S|M,L) = 1.04276977	2.68	0.0853	2	0.50903840	1.31	0.2854
STRM_LOC*SWMP_COM	2	R(I|M,L,S) = 0.60773900	1.56	0.2266	2	0.60773900	1.56	0.2266

PARAMETER			ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		$\bar{y}_{13}$ =	6.12000000 B	31.04	0.0001	0.19715797
STRM_LOC	A	$\bar{y}_{23}-\bar{y}_{13}$ =	0.78000000 B	2.92	0.0067	0.26695315
	N		0.00000000 B	.	.	.
SWMP_COM	M	$\bar{y}_{12}-\bar{y}_{13}$ =	0.34666667 B	1.30	0.2043	0.26695315
	N	$\bar{y}_{11}-\bar{y}_{13}$ =	0.70500000 B	2.81	0.0089	0.25132808
	S		0.00000000 B	.	.	.
STRM_LOC*SWMP_COM	A M	$\bar{y}_{22}-\bar{y}_{23}-\bar{y}_{12}+\bar{y}_{13}$ =	-0.49666667 B	-1.39	0.1756	0.35770273
	A N	$\bar{y}_{21}-\bar{y}_{23}-\bar{y}_{11}+\bar{y}_{13}$ =	-0.70500000 B	-1.61	0.1191	0.43901787
	A S		0.00000000 B	.	.	.
	N M		0.00000000 B	.	.	.
	N N		0.00000000 B	.	.	.
	N S		0.00000000 B	.	.	.

note that the estimates are not identical to those on (F) because SAS orders the levels alphabetically

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

(B)

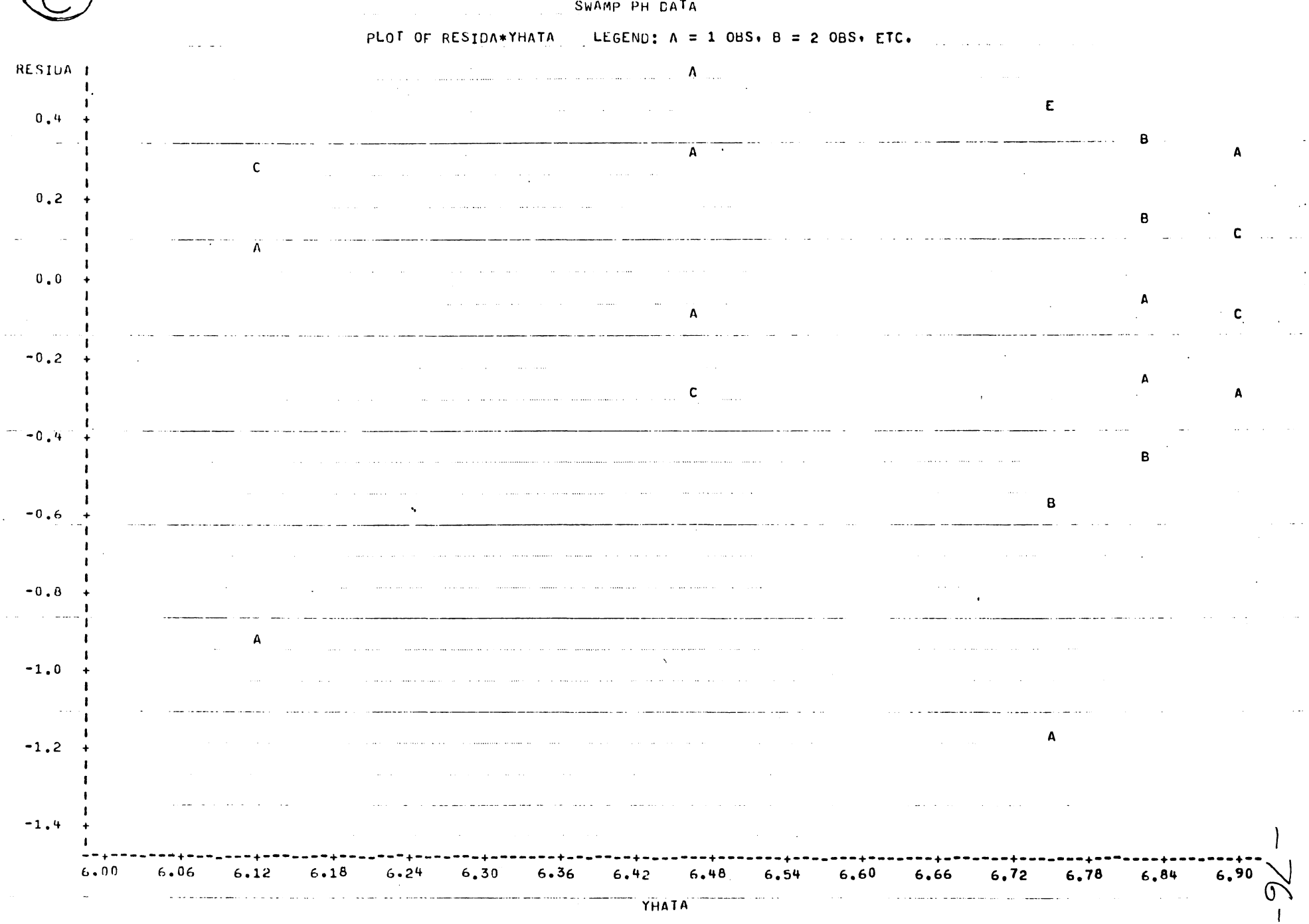
For models in (A), (D), (F) & (H)

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	6.60000000	6.82500000	-0.22500000
2	7.20000000	6.82500000	0.37500000
3	7.20000000	6.82500000	0.37500000
4	7.00000000	6.82500000	0.17500000
5	6.80000000	6.82500000	-0.02500000
6	6.40000000	6.82500000	-0.42500000
7	7.00000000	6.82500000	0.17500000
8	6.40000000	6.82500000	-0.42500000
9	6.80000000	6.46666667	0.33333333
10	7.00000000	6.46666667	0.53333333
11	6.20000000	6.46666667	-0.26666667
12	6.20000000	6.46666667	-0.26666667
13	6.40000000	6.46666667	-0.06666667
14	6.20000000	6.46666667	-0.26666667
15	6.40000000	6.12000000	0.28000000
16	5.20000000	6.12000000	-0.92000000
17	6.20000000	6.12000000	0.08000000
18	6.40000000	6.12000000	0.28000000
19	6.40000000	6.12000000	0.28000000
20	6.80000000	6.90000000	-0.10000000
21	7.00000000	6.90000000	0.10000000
22	6.20000000	6.75000000	-0.55000000
23	5.60000000	6.75000000	-1.15000000
24	7.20000000	6.75000000	0.45000000
25	7.20000000	6.75000000	0.45000000
26	7.20000000	6.75000000	0.45000000
27	6.20000000	6.75000000	-0.55000000
28	7.20000000	6.75000000	0.45000000
29	7.20000000	6.75000000	0.45000000
30	7.20000000	6.90000000	0.30000000
31	6.80000000	6.90000000	-0.10000000
32	7.00000000	6.90000000	0.10000000
33	6.80000000	6.90000000	-0.10000000
34	7.00000000	6.90000000	0.10000000
35	6.60000000	6.90000000	-0.30000000

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	5.63633333
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.02583860
DURBIN-WATSON D	2.02672788

(C)

Residual plot for models in (A), (D), (F) & (H).



SAS orders
factors
alphabetically
and levels within
a factor
alphabetically

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR COLUMN DIFFERENCE 1

EFFECT		COEFFICIENTS
INTERCEPT		0
SWMP_COM	M	0
	N	1
	S	-1
STRM_LOC	A	0
	N	0
STRM_LOC*SWMP_COM	A M	0
	A N	0.5
	A S	-0.5
	N M	0
	N N	0.5
	N S	-0.5

DEPENDENT VARIABLE: Y

ESTIMABLE FUNCTIONS FOR COLUMN DIFFERENCE 2

EFFECT		COEFFICIENTS
INTERCEPT		0
SWMP_COM	M	1
	N	-0.5
	S	-0.5
STRM_LOC	A	0
	N	0
STRM_LOC*SWMP_COM	A M	0.5
	A N	-0.25
	A S	-0.25
	N M	0.5
	N N	-0.25
	N S	-0.25

(D)

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	2.42538095	0.48507619	2.50	0.0536	0.300852	6.6167
ERROR	29	5.63633333	0.19435632		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429			0.44085862		6.66285714

ANOVA(2) of p.18.16

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
SWMP_COM	2	0.46187013	1.24	0.3044	2	0.50903840	1.31	0.2854
STRM_LOC	1	1.33577182	6.87	0.0138	1	1.00971645	5.20	0.0302
STRM_LOC*SWMP_COM	2	0.60773900	1.56	0.2266	2	0.60773900	1.56	0.2266

PARAMETER			ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		$\bar{y}_{13}$ =	6.12000000 B	31.04	0.0001	0.19715797
SWMP_COM	M	$\bar{y}_{12}-\bar{y}_{13}$ =	0.34666667 B	1.30	0.2043	0.26695315
	N	$\bar{y}_{11}-\bar{y}_{13}$ =	0.70500000 B	2.81	0.0089	0.25132808
	S		0.00000000 B	.	.	.
STRM_LOC	A	$\bar{y}_{23}-\bar{y}_{13}$ =	0.78000000 B	2.92	0.0067	0.26695315
	N		0.00000000 B	.	.	.
STRM_LOC*SWMP_COM	A M	$\bar{y}_{22}-\bar{y}_{23}-\bar{y}_{12}+\bar{y}_{13}$ =	-0.49666667 B	-1.39	0.1756	0.35770273
	A N	$\bar{y}_{21}-\bar{y}_{23}-\bar{y}_{11}+\bar{y}_{13}$ =	-0.70500000 B	-1.61	0.1191	0.43901787
	A S		0.00000000 B	.	.	.
	N M		0.00000000 B	.	.	.
	N N		0.00000000 B	.	.	.
	N S		0.00000000 B	.	.	.

note that the estimates are different from those on (H), same reason as on (A)

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
ROW DIFFERENCES	0.37944444	2.28	0.0302	0.16647454
COLUMN DIFFERENCE 1	0.35250000	1.61	0.1191	0.21950893
COLUMN DIFFERENCE 2	-0.07791667	-0.48	0.6340	0.16191929

These are natural contrasts

(E)

SWAMP PH DATA

BS	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	1	1	0	1	0	0	1	0	0	0	0	0
2	1	1	0	1	0	0	1	0	0	0	0	0
3	1	1	0	1	0	0	1	0	0	0	0	0
4	1	1	0	1	0	0	1	0	0	0	0	0
5	1	1	0	1	0	0	1	0	0	0	0	0
6	1	1	0	1	0	0	1	0	0	0	0	0
7	1	1	0	1	0	0	1	0	0	0	0	0
8	1	1	0	1	0	0	1	0	0	0	0	0
9	1	1	0	0	1	0	0	1	0	0	0	0
10	1	1	0	0	1	0	0	1	0	0	0	0
11	1	1	0	0	1	0	0	1	0	0	0	0
12	1	1	0	0	1	0	0	1	0	0	0	0
13	1	1	0	0	1	0	0	1	0	0	0	0
14	1	1	0	0	1	0	0	1	0	0	0	0
15	1	1	0	0	0	1	0	0	1	0	0	0
16	1	1	0	0	0	1	0	0	1	0	0	0
17	1	1	0	0	0	1	0	0	1	0	0	0
18	1	1	0	0	0	1	0	0	1	0	0	0
19	1	1	0	0	0	1	0	0	1	0	0	0
20	1	0	1	1	0	0	0	0	0	1	0	0
21	1	0	1	1	0	0	0	0	0	1	0	0
22	1	0	1	0	1	0	0	0	0	0	1	0
23	1	0	1	0	1	0	0	0	0	0	1	0
24	1	0	1	0	1	0	0	0	0	0	1	0
25	1	0	1	0	1	0	0	0	0	0	1	0
26	1	0	1	0	1	0	0	0	0	0	1	0
27	1	0	1	0	1	0	0	0	0	0	1	0
28	1	0	1	0	1	0	0	0	0	0	1	0
29	1	0	1	0	1	0	0	0	0	0	1	0
30	1	0	1	0	0	1	0	0	0	0	0	1
31	1	0	1	0	0	1	0	0	0	0	0	1
32	1	0	1	0	0	1	0	0	0	0	0	1
33	1	0	1	0	0	1	0	0	0	0	0	1
34	1	0	1	0	0	1	0	0	0	0	0	1
35	1	0	1	0	0	1	0	0	0	0	0	1

X matrix for model in (F)

(F)

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	2.42538095	0.48507619	2.50	0.0536	0.300852	6.6167
ERROR	29	residual: 5.63633333	0.19435632		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429 = 1561.85 − 1553.79			0.44085862		6.66285714

$R(M) = 35 \times \bar{Y}^2 = 1553.79$

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	R(LIM)= 0.77487218	3.99	0.0553	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.
X3	1	R(SIM,L)={1.01055615} = 1.0428	5.20	0.0301	0	0.00000000	.	.
X4	1	{0.03221652}	0.17	0.6869	0	0.00000000	.	.
X5	0	0.00000000	.	.	0	0.00000000	.	.
X6	1	R(I|M,L,S)={0.23303858} = .6077	1.20	0.2825	0	0.00000000	.	.
X7	1	{0.37470042}	1.93	0.1756	0	0.00000000	.	.
X8	0	0.00000000	.	.	0	0.00000000	.	.
X9	0	0.00000000	.	.	0	0.00000000	.	.
X10	0	0.00000000	.	.	0	0.00000000	.	.
X11	0	0.00000000	.	.	0	0.00000000	.	.

↑ ANOVA (1) on p. 18.16

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	$\bar{y}_{23}$ =	6.90000000 B	38.34	0.0001	0.17997978
X1	$\bar{y}_{13} - \bar{y}_{23}$ =	-0.78000000 B	-2.92	0.0067	0.26695315
X2		0.00000000 B	.	.	.
X3	$\bar{y}_{21} - \bar{y}_{23}$ =	-0.00000000 B	-0.00	1.0000	0.35995955
X4	$\bar{y}_{22} - \bar{y}_{23}$ =	-0.15000000 B	-0.63	0.5336	0.23809087
X5		0.00000000 B	.	.	.
X6	$\bar{y}_{11} - \bar{y}_{13} - \bar{y}_{21} + \bar{y}_{23}$ =	0.70500000 B	1.61	0.1191	0.43901787
X7	$\bar{y}_{12} - \bar{y}_{13} - \bar{y}_{22} + \bar{y}_{23}$ =	0.49666667 B	1.39	0.1756	0.35770273
X8		0.00000000 B	.	.	.
X9		0.00000000 B	.	.	.
X10		0.00000000 B	.	.	.
X11		0.00000000 B	.	.	.

↑ compare with p. 18.15

model:
equal means | row indicators | column indicators | general means

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

(G)

SWAMP PH DATA

OBS	X0	X3	X4	X5	X1	X2	X6	X7	X8	X9	X10	X11
1	1	1	0	0	1	0	1	0	0	0	0	0
2	1	1	0	0	1	0	1	0	0	0	0	0
3	1	1	0	0	1	0	1	0	0	0	0	0
4	1	1	0	0	1	0	1	0	0	0	0	0
5	1	1	0	0	1	0	1	0	0	0	0	0
6	1	1	0	0	1	0	1	0	0	0	0	0
7	1	1	0	0	1	0	1	0	0	0	0	0
8	1	1	0	0	1	0	1	0	0	0	0	0
9	1	0	1	0	1	0	0	1	0	0	0	0
10	1	0	1	0	1	0	0	1	0	0	0	0
11	1	0	1	0	1	0	0	1	0	0	0	0
12	1	0	1	0	1	0	0	1	0	0	0	0
13	1	0	1	0	1	0	0	1	0	0	0	0
14	1	0	1	0	1	0	0	1	0	0	0	0
15	1	0	0	1	1	0	0	0	1	0	0	0
16	1	0	0	1	1	0	0	0	1	0	0	0
17	1	0	0	1	1	0	0	0	1	0	0	0
18	1	0	0	1	1	0	0	0	1	0	0	0
19	1	0	0	1	1	0	0	0	1	0	0	0
20	1	1	0	0	0	1	0	0	0	1	0	0
21	1	1	0	0	0	1	0	0	0	1	0	0
22	1	0	1	0	0	1	0	0	0	0	1	0
23	1	0	1	0	0	1	0	0	0	0	1	0
24	1	0	1	0	0	1	0	0	0	0	1	0
25	1	0	1	0	0	1	0	0	0	0	1	0
26	1	0	1	0	0	1	0	0	0	0	1	0
27	1	0	1	0	0	1	0	0	0	0	1	0
28	1	0	1	0	0	1	0	0	0	0	1	0
29	1	0	1	0	0	1	0	0	0	0	1	0
30	1	0	0	1	0	1	0	0	0	0	0	1
31	1	0	0	1	0	1	0	0	0	0	0	1
32	1	0	0	1	0	1	0	0	0	0	0	1
33	1	0	0	1	0	1	0	0	0	0	0	1
34	1	0	0	1	0	1	0	0	0	0	0	1
35	1	0	0	1	0	1	0	0	0	0	0	1

X-matrix for model in (H)

(H)

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	2.42538095	0.48507619	2.50	0.0536	0.300852	6.6167
ERROR	29	Residual = 5.63633333	0.19435632		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429			0.44085862		6.66285714

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X3	1	R(SIM) = {0.43931429} = .4819	2.26	0.1435	0	0.00000000	.	.
X4	1	{0.04255584}	0.22	0.6433	0	0.00000000	.	.
X5	0	0.00000000	.	.	0	0.00000000	.	.
X1	1	R(LIM,S) = 1.33577182	6.87	0.0138	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.
X6	1	R(LIM,S,L) = {0.23303858} = .6077	1.20	0.2825	0	0.00000000	.	.
X7	1	{0.37470042}	1.93	0.1756	0	0.00000000	.	.
X8	0	0.00000000	.	.	0	0.00000000	.	.
X9	0	0.00000000	.	.	0	0.00000000	.	.
X10	0	0.00000000	.	.	0	0.00000000	.	.
X11	0	0.00000000	.	.	0	0.00000000	.	.

↑ ANOVA (2) on p. 18.16

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	$\bar{y}_{23}$ =	6.90000000 B	38.34	0.0001	0.17997978
X3	$\bar{y}_{21} - \bar{y}_{23}$ =	-0.00000000 B	-0.00	1.0000	0.35995955
X4	$\bar{y}_{22} - \bar{y}_{23}$ =	-0.15000000 B	-0.63	0.5336	0.23809087
X5		0.00000000 B	.	.	.
X1	$\bar{y}_{13} - \bar{y}_{23}$ =	-0.78000000 B	-2.92	0.0067	0.26695315
X2		0.00000000 B	.	.	.
X6	$\bar{y}_{11} - \bar{y}_{13} - \bar{y}_{21} + \bar{y}_{23}$ =	0.70500000 B	1.61	0.1191	0.43901787
X7	$\bar{y}_{12} - \bar{y}_{13} - \bar{y}_{22} + \bar{y}_{23}$ =	0.49666667 B	1.39	0.1756	0.35770273
X8		0.00000000 B	.	.	.
X9		0.00000000 B	.	.	.
X10		0.00000000 B	.	.	.
X11		0.00000000 B	.	.	.

↑ see p. 18.15

Model:
equal means | column indicators | row indicators | general means |

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

(I)

SWAMP PH DATA

OBS	X0	X1	X2	X3	X4	X5
1	1	1	0	1	0	0
2	1	1	0	1	0	0
3	1	1	0	1	0	0
4	1	1	0	1	0	0
5	1	1	0	1	0	0
6	1	1	0	1	0	0
7	1	1	0	1	0	0
8	1	1	0	1	0	0
9	1	1	0	0	1	0
10	1	1	0	0	1	0
11	1	1	0	0	1	0
12	1	1	0	0	1	0
13	1	1	0	0	1	0
14	1	1	0	0	1	0
15	1	1	0	0	0	1
16	1	1	0	0	0	1
17	1	1	0	0	0	1
18	1	1	0	0	0	1
19	1	1	0	0	0	1
20	1	0	1	1	0	0
21	1	0	1	1	0	0
22	1	0	1	0	1	0
23	1	0	1	0	1	0
24	1	0	1	0	1	0
25	1	0	1	0	1	0
26	1	0	1	0	1	0
27	1	0	1	0	1	0
28	1	0	1	0	1	0
29	1	0	1	0	1	0
30	1	0	1	0	0	1
31	1	0	1	0	0	1
32	1	0	1	0	0	1
33	1	0	1	0	0	1
34	1	0	1	0	0	1
35	1	0	1	0	0	1

X-matrix for a reduced model (J) of the model in (F)

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	6.60000000	6.75699933	-0.15699933
2	7.20000000	6.75699933	0.44300067
3	7.20000000	6.75699933	0.44300067
4	7.00000000	6.75699933	0.24300067
5	6.80000000	6.75699933	0.04300067
6	6.40000000	6.75699933	-0.35699933
7	7.00000000	6.75699933	0.24300067
8	6.40000000	6.75699933	-0.35699933
9	6.80000000	6.39142666	0.40857334
10	7.00000000	6.39142666	0.60857334
11	6.20000000	6.39142666	-0.19142666
12	6.20000000	6.39142666	-0.19142666
13	6.40000000	6.39142666	0.00857334
14	6.20000000	6.39142666	-0.19142666
15	6.40000000	6.31908908	0.08091092
16	5.20000000	6.31908908	-1.11908908
17	6.20000000	6.31908908	-0.11908908
18	6.40000000	6.31908908	0.08091092
19	6.40000000	6.31908908	0.08091092
20	6.80000000	7.17200268	-0.37200268
21	7.00000000	7.17200268	-0.17200268
22	6.20000000	6.80643001	-0.60643001
23	5.60000000	6.80643001	-1.20643001
24	7.20000000	6.80643001	0.39356999
25	7.20000000	6.80643001	0.39356999
26	7.20000000	6.80643001	0.39356999
27	6.20000000	6.80643001	-0.60643001
28	7.20000000	6.80643001	0.39356999
29	7.20000000	6.80643001	0.39356999
30	7.20000000	6.73409243	0.46590757
31	6.80000000	6.73409243	0.06590757
32	7.00000000	6.73409243	0.26590757
33	6.80000000	6.73409243	0.06590757
34	7.00000000	6.73409243	0.26590757
35	6.60000000	6.73409243	-0.13409243

STATISTICAL ANALYSIS SYSTEM

17:12 SUNDAY

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	6.24407234
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	0.11420073
DURBIN-WATSON D	1.76477133

J

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	1.81764195	0.60588065	3.01	0.0451	0.225466	6.7359
ERROR	31	6.24407234	0.20142169		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429			0.44880028		6.66285714

ANOVA to the left on bottom of p. 18.18

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	R(L|M)= 0.77487218	3.85	0.0589	0	0.00000000	.	.
X2	0	0.00000000	.	.	0	0.00000000	.	.
X3	1	R(S|M,L)={1.01055345}=1.043	5.02 } 2.59	0.0324 ns	0	0.00000000	.	.
X4	1	{0.03221632}	0.16	0.6919	0	0.00000000	.	.
X5	0	0.00000000	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	6.73409243 B	43.76	0.0001	0.15387277
X1	-0.41500335 B	-2.58	0.0150	0.16115307
X2	0.00000000 B	.	.	.
X3	0.43791025 B	2.15	0.0396	0.20384430
X4	0.07233758 B	0.40	0.6919	0.18087522
X5	0.00000000 B	.	.	.

Reduced model: equal means / row indicators / column indicators

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

Rows / Columns is N-N (at 5% level) (see p. 18.19)

(K)

SWAMP PH DATA

OBS	X0	X3	X4	X5	X1	X2
1	1	1	0	0	1	0
2	1	1	0	0	1	0
3	1	1	0	0	1	0
4	1	1	0	0	1	0
5	1	1	0	0	1	0
6	1	1	0	0	1	0
7	1	1	0	0	1	0
8	1	1	0	0	1	0
9	1	0	1	0	1	0
10	1	0	1	0	1	0
11	1	0	1	0	1	0
12	1	0	1	0	1	0
13	1	0	1	0	1	0
14	1	0	1	0	1	0
15	1	0	0	1	1	0
16	1	0	0	1	1	0
17	1	0	0	1	1	0
18	1	0	0	1	1	0
19	1	0	0	1	1	0
20	1	1	0	0	0	1
21	1	1	0	0	0	1
22	1	0	1	0	0	1
23	1	0	1	0	0	1
24	1	0	1	0	0	1
25	1	0	1	0	0	1
26	1	0	1	0	0	1
27	1	0	1	0	0	1
28	1	0	1	0	0	1
29	1	0	1	0	0	1
30	1	0	0	1	0	1
31	1	0	0	1	0	1
32	1	0	0	1	0	1
33	1	0	0	1	0	1
34	1	0	0	1	0	1
35	1	0	0	1	0	1

X-matrix for model in (L), a reduced model of that in (H)

(L)

SWAMP PH DATA

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	1.81764195	0.60588065	3.01	0.0451	0.225466	6.7359
ERROR	31	6.24407234	0.20142169		STD DEV		Y MEAN
CORRECTED TOTAL	34	8.06171429			0.44880028		6.66285714

ANOVA to the right on bottom of p. 18.18

SOURCE	DF	TYPE I SS	F VALUE	PR > F		DF	TYPE IV SS	F VALUE	PR > F
X3	1	R(S/M) = {0.43931429} = .4819	2.18	0.1498	ns	0	0.00000000	.	.
X4	1	{0.04255584}	0.21	0.6490		0	0.00000000	.	.
X5	0	0.00000000	.	.		0	0.00000000	.	.
X1	1	R(L/M,S) = 1.33577182	6.63	0.0150		0	0.00000000	.	.
X2	0	0.00000000	.	.		0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	6.73409243 B	43.76	0.0001	0.15387277
X3	0.43791025 B	2.15	0.0396	0.20384430
X4	0.07233758 B	0.40	0.6919	0.18087522
X5	0.00000000 B	.	.	.
X1	-0.41500335 B	-2.58	0.0150	0.16115307
X2	0.00000000 B	.	.	.

Reduced model: equal means | column ind's row indicators

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

Columns | Rows is N - I (see p. 18.19) (at 5% level)
Reading off table p. 18.19 we find C|R which means that only the sequence columns | rows needs to be fitted.

1 SOYBEAN PHYSIOLOGICAL DATA

NOTE: THE JOB VMPJQ0C1 HAS BEEN RUN UNDER RELEASE 79.3 OF SAS AT CORNELL UNIVERS

```
1          DATA SOYBEAN;
2               INPUT OBS LIGHT $ HEIGHT YIELD;
3               Y=YIELD;
4               X4=HEIGHT;
5               X1=(LIGHT = 'C');
6               X2=(LIGHT = 'L');
7               X3=1-X1-X2;
8               X5=X1*X4;
9               X6=X2*X4;
10              X7=X3*X4;
11         CARDS;
```

UNIT 19

covariate

treatment indicators

heights for each treatment

see p 19-8

See UNIT 19 for the data

NOTE: DATA SET WORK.SOYBEAN HAS 45 OBSERVATIONS AND 12 VARIABLES. 130 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.26 SECONDS AND 128K.

```
57         PROC GLM;
58              CLASS LIGHT;
59              MODEL Y=LIGHT/SOLUTION P;
60              OUTPUT OUT=NEW1      RESIDUAL=RESID1      PREDICTED=YHAT1;
```

(A)

Produces same ANOVA as on p. 19-2.

NOTE: DATA SET WORK.NEW1 HAS 45 OBSERVATIONS AND 14 VARIABLES. 112 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.44 SECONDS AND 178K AND PRINTED PAGES 1 TO 3.

```
61         PROC PLOT;
62              PLOT RESID1*YHAT1;
```

(B) Residual plot for model in (A)

NOTE: THE PROCEDURE PLOT USED 0.23 SECONDS AND 130K AND PRINTED PAGE 4.

```
63         PROC GLM;
64              MODEL Y=X1 X2 X3 X4/NOINT P;
65              OUTPUT OUT=NEW2      RESIDUAL=RESID2      PREDICTED=YHAT2;
```

(C)

Model: treatments then covariate

NOTE: DATA SET WORK.NEW2 HAS 45 OBSERVATIONS AND 16 VARIABLES. 98 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.44 SECONDS AND 180K AND PRINTED PAGES 5 TO 6.

```
66         PROC PLOT;
67              PLOT Y*X4 RESID2*YHAT2;
```

(D) (E)

Plot of y versus covariate and residual plot for model in (C) and (F)

NOTE: THE PROCEDURE PLOT USED 0.32 SECONDS AND 138K AND PRINTED PAGES 7 TO 8.

```
68         PROC GLM;
69              MODEL Y=X4 X1-X3/P;
```

(F)

Model: covariate then treatments

NOTE: THE PROCEDURE GLM USED 0.41 SECONDS AND 160K AND PRINTED PAGES 9 TO 10.

```
70         PROC GLM;
71              MODEL Y=X1-X7/P;
72              OUTPUT OUT=NEW3      RESIDUAL=RESID3      PREDICTED=YHAT3;
```

(G)

Model: equal means / treatments / common slope / different slopes

NOTE: DATA SET WORK.NEW3 HAS 45 OBSERVATIONS AND 18 VARIABLES. 88 OBS/TRK.
NOTE: THE PROCEDURE GLM USED 0.51 SECONDS AND 180K AND PRINTED PAGES 11 TO 12.

```
73         PROC GLM;
74              MODEL Y=X4 X1-X3 X5-X7/P;
```

(H)

Using the approach of p. 19-12

model: equal means / common slope / different intercepts / different slopes

NOTE: THE PROCEDURE GLM USED 0.47 SECONDS AND 164K AND PRINTED PAGES 13 TO 14.

75 (I) PROC PLOT;
76 PLOT RESID3*YHAT3;

Residual plot for models in (G) & (H)

NOTE: THE PROCEDURE PLOT USED 0.24 SECONDS AND 130K AND PRINTED PAGE 15.

X-matrices :

(A) (C) (G)

X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	0	0	0	40	40	0	0
1	0	0	0	61	61	0	0
1	0	0	0	56	56	0	0
1	0	0	0	42	42	0	0
1	0	0	0	50	50	0	0
1	0	0	0	51	51	0	0
1	0	0	0	65	65	0	0
1	0	0	0	35	35	0	0
1	0	0	0	52	52	0	0
1	0	0	0	48	48	0	0
1	0	0	0	60	60	0	0
1	0	0	0	33	33	0	0
1	0	0	0	51	51	0	0
1	0	0	0	48	48	0	0
1	0	0	0	48	48	0	0
1	0	0	0	33	0	33	0
1	0	0	0	63	0	63	0
1	0	0	0	63	0	63	0
1	0	0	0	35	0	35	0
1	0	0	0	50	0	50	0
1	0	0	0	38	0	38	0
1	0	0	0	50	0	50	0
1	0	0	0	49	0	49	0
1	0	0	0	50	0	50	0
1	0	0	0	45	0	45	0
1	0	0	0	49	0	49	0
1	0	0	0	50	0	50	0
1	0	0	0	62	0	62	0
1	0	0	0	62	0	62	0
1	0	0	0	48	0	48	0
1	0	0	1	52	0	0	52
1	0	0	1	57	0	0	57
1	0	0	1	55	0	0	55
1	0	0	1	55	0	0	55
1	0	0	1	54	0	0	54
1	0	0	1	66	0	0	66
1	0	0	1	45	0	0	45
1	0	0	1	67	0	0	67
1	0	0	1	41	0	0	41
1	0	0	1	67	0	0	67
1	0	0	1	40	0	0	40
1	0	0	1	58	0	0	58
1	0	0	1	56	0	0	56
1	0	0	1	62	0	0	62
1	0	0	1	52	0	0	52

(F) is: $X_0\ X_4\ X_1 X_2 X_3$

(H) is $X_0\ X_4\ X_1 X_2 X_3 X_5 X_6 X_7$

STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
LIGHT	3	C L S

NUMBER OF OBSERVATIONS IN DATA SET = 45

(A)

Means model

$y_{ij} = \mu_j + \varepsilon_{ij}$

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	2	301.83793778	150.91896889	490.56	0.0001	0.958949	4.4173
ERROR	42	12.92109333	0.30764508		STD DEV		Y MEAN
CORRECTED TOTAL	44	314.75903111			0.55465762		12.55644444
uncorrected total	45	7409.662396					R(0) = 7094.893365

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
LIGHT	2	301.83793778	490.56	0.0001	2	301.83793778	490.56	0.0001

Compare with ANOVA p. 19-2

PARAMETER			ESTIMATE		T FOR H0: PARAMETER=0	PR > ITI	STD ERROR OF ESTIMATE
INTERCEPT		$\bar{y}_3 =$	9.50933333	B	66.40	0.0001	0.14321198
LIGHT	C	$\bar{y}_1 - \bar{y}_3 =$	2.91066667	B	13.88	0.0001	0.20253233
	L	$\bar{y}_2 - \bar{y}_3 =$	6.33066667	B	31.26	0.0001	0.20253233
	S		0.00000000	B	.	.	.

$\bar{y}_1 = 12.32000000$ $\bar{y}_2 = 15.84000000$

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: (BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

10

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL	
1	11.88000000	12.32000000	-0.44000000	
2	12.76000000	12.32000000	0.44000000	
3	12.68000000	12.32000000	0.36000000	
4	11.96000000	12.32000000	-0.36000000	
5	12.40000000	12.32000000	0.08000000	
6	12.24000000	12.32000000	-0.08000000	
7	13.25000000	12.32000000	0.93000000	←
8	11.39000000	12.32000000	-0.93000000	←
9	12.45000000	12.32000000	0.13000000	
10	12.19000000	12.32000000	-0.13000000	
11	13.18000000	12.32000000	0.86000000	←
12	11.46000000	12.32000000	-0.86000000	←
13	12.37000000	12.32000000	0.05000000	
14	12.27000000	12.32000000	-0.05000000	
15	12.32000000	12.32000000	0.00000000	
16	15.00000000	15.84000000	-0.84000000	←
17	16.68000000	15.84000000	0.84000000	←
18	16.59000000	15.84000000	0.75000000	
19	15.09000000	15.84000000	-0.75000000	
20	16.24000000	15.84000000	0.40000000	
21	15.44000000	15.84000000	-0.40000000	
22	15.80000000	15.84000000	-0.04000000	
23	15.88000000	15.84000000	0.04000000	
24	16.00000000	15.84000000	0.16000000	
25	15.68000000	15.84000000	-0.16000000	
26	15.82000000	15.84000000	-0.02000000	
27	15.86000000	15.84000000	0.02000000	
28	14.98000000	15.84000000	-0.86000000	←
29	16.70000000	15.84000000	0.86000000	←
30	15.84000000	15.84000000	0.00000000	
31	9.50000000	9.50933333	-0.00933333	
32	9.52000000	9.50933333	0.01066667	
33	9.51000000	9.50933333	0.00066667	
34	9.45000000	9.50933333	-0.05933333	
35	9.56000000	9.50933333	0.05066667	
36	10.20000000	9.50933333	0.69066667	
37	8.82000000	9.50933333	-0.68933333	
38	10.37000000	9.50933333	0.86066667	←
39	8.65000000	9.50933333	-0.85933333	←
40	10.49000000	9.50933333	0.98066667	←
41	8.53000000	9.50933333	-0.97933333	←
42	9.64000000	9.50933333	0.13066667	
43	9.38000000	9.50933333	-0.12933333	
44	9.85000000	9.50933333	0.34066667	
45	9.17000000	9.50933333	-0.33933333	

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	12.92109333
SUM OF SQUARED RESIDUALS - ERROR SS	-0.00000000
FIRST ORDER AUTOCORRELATION	-0.61261641
DURBIN-WATSON D	3.20133300

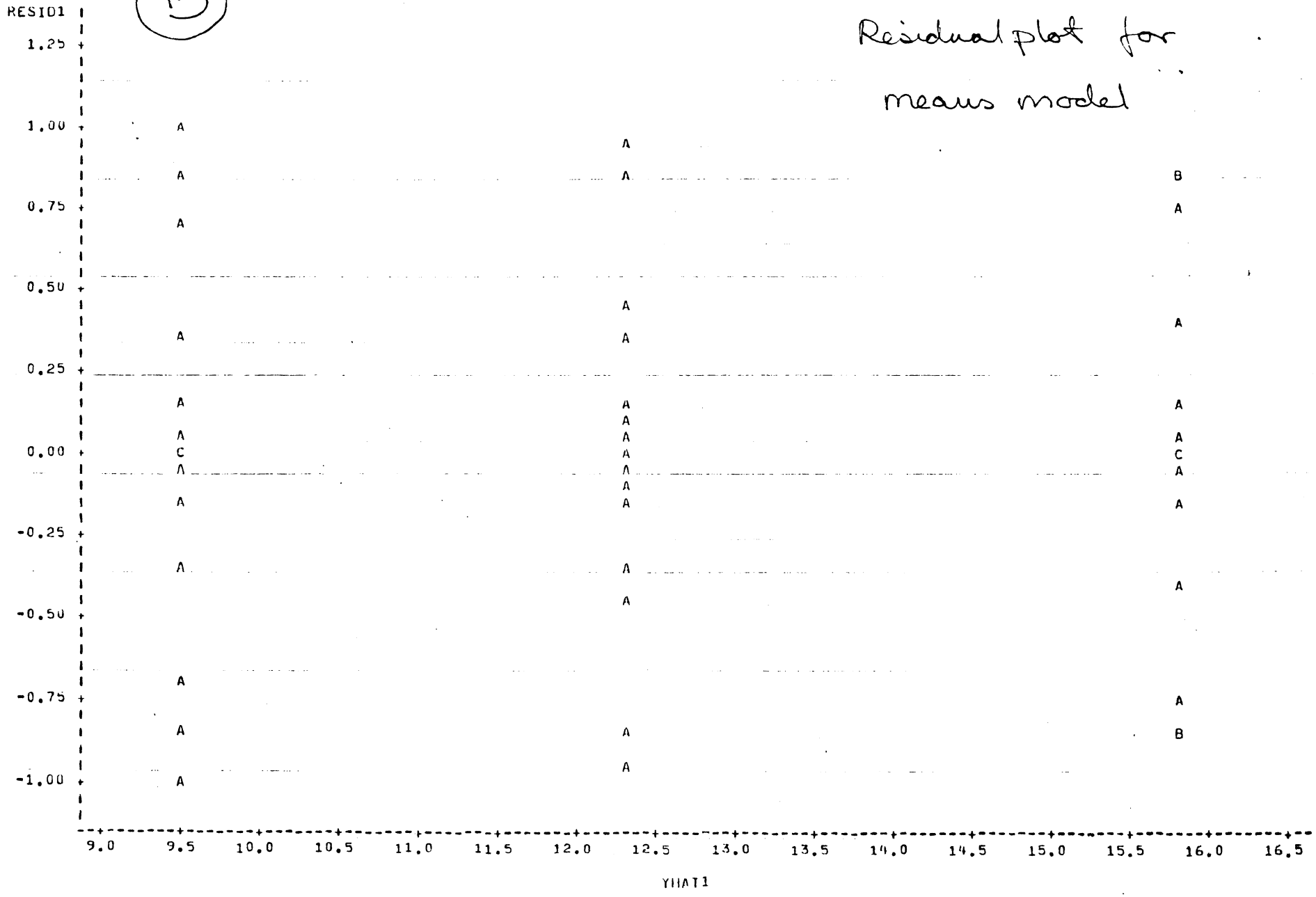
← means rather large residuals, compare with residuals on p. 95

S T A T I S T I C A L A N A L Y S I S S Y S T E M

PLOT OF RESID1*YHAT1 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

(B)

Residual plot for means model



STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

Means model with covariate

$y_{ik} = \beta_i + \beta x_{ik} + \varepsilon_{ik}$

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	7407.90044470	1851.97511118	43340.71	0.0001	0.999764	1.6463
ERROR	41	1.75195530	0.04273062		STD DEV		Y MEAN
UNCORRECTED TOTAL	45	7409.65240000			0.20671385		12.55644444

compare this residual to the RSS of the means model

see discussion on p 19-4

SOURCE	DF	TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	2276.73600000	53281.14	0.0001	1	112.96933203	2643.76	0.0001
X2	1	3763.50400000	88076.99	0.0001	1	214.52633803	5020.44	0.0001
X3	1	1356.41130667	31745.31	0.0001	1	41.02687687	960.13	0.0001
X4	1	11.16913803	261.38	0.0001	1	11.16913803	261.38	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
X1 intercept for tr1 =	9.47006202	51.42	0.0001	0.18417985
X2 intercept for tr2 =	13.00161502	70.86	0.0001	0.18349600
X3 intercept for tr3 =	6.32433507	30.99	0.0001	0.20410339
X4 b =	0.05776901	16.17	0.0001	0.00357318

$\rightarrow$ estimates + $b * \bar{x}$ = .05776901 * 51.20 + estimates

= estimates + 2.9577733312 = $\begin{cases} 12.42783533 = \bar{y}_1 \text{ (adj)} \\ 15.95938913 = \bar{y}_2 \text{ (adj)} \\ 9.28210838 = \bar{y}_3 \text{ (adj)} \end{cases}$

compare with p 19-5

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL	
1	11.88000000	11.78082254	0.09917746	
2	12.76000000	12.99397182	-0.23397182	
3	12.68000000	12.70512675	-0.02512675	
4	11.96000000	11.89636057	0.06363943	
5	12.40000000	12.35851268	0.04148732	
6	12.24000000	12.41628169	-0.17628169	
7	13.25000000	13.22504787	0.02495213	
8	11.39000000	11.49197748	-0.10197748	
9	12.45000000	12.47405070	-0.02405070	
10	12.19000000	12.24297465	-0.05297465	
11	13.18000000	12.93620281	0.24379719	
12	11.46000000	11.37643945	0.08356055	
13	12.37000000	12.41628169	-0.04628169	
14	12.27000000	12.24297465	0.02702535	
15	12.32000000	12.24297465	0.07702535	
16	15.00000000	14.90799325	0.09200675	
17	16.68000000	16.64106365	0.03893635	
18	16.59000000	16.64106365	-0.05106365	
19	15.09000000	15.02353128	0.06646872	
20	16.24000000	15.89006648	0.34993352	
21	15.44000000	15.19683832	0.24316168	
22	15.80000000	15.89006648	-0.09006648	
23	15.88000000	15.83229746	0.04770254	
24	16.00000000	15.89006648	0.10993352	
25	15.68000000	15.60122141	0.07877859	
26	15.82000000	15.83229746	-0.01229746	
27	15.86000000	15.89006648	-0.03006648	
28	14.98000000	16.00560450	-1.02560450	← only large residual
29	16.70000000	16.58329464	0.11670536	
30	15.84000000	15.77452845	0.06547155	
31	9.50000000	9.32832376	0.17167624	
32	9.52000000	9.61716882	-0.09716882	
33	9.51000000	9.50163080	0.00836920	
34	9.45000000	9.50163080	-0.05163080	
35	9.56000000	9.44386179	0.11613821	
36	10.20000000	10.13708994	0.06291006	
37	8.82000000	8.92394067	-0.10394067	
38	10.37000000	10.19485896	0.17514104	
39	8.65000000	8.69286461	-0.04286461	
40	10.49000000	10.19485896	0.29514104	
41	8.53000000	8.63509560	-0.10509560	
42	9.64000000	9.67493784	-0.03493784	
43	9.38000000	9.55939981	-0.17939981	
44	9.85000000	9.90601389	-0.05601389	
45	9.17000000	9.32832376	-0.15832376	

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	1.75195530
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.03156765
DURBIN-WATSON D	2.04321323

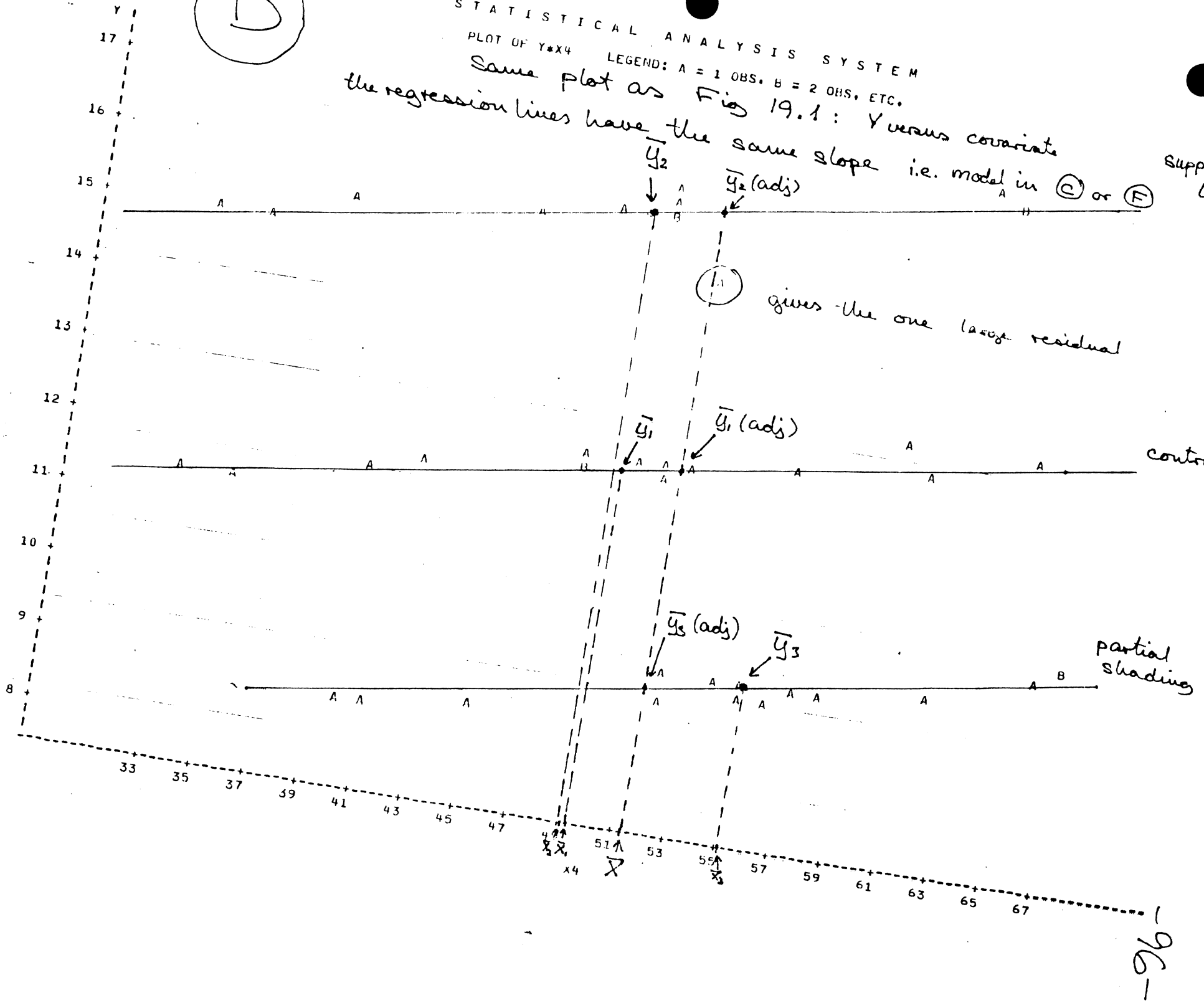
(D)

STATISTICAL ANALYSIS SYSTEM

PLOT OF Y*X4 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Same plot as Fig 19.1 : Y versus covariate

the regression lines have the same slope i.e. model in (C) or (E)

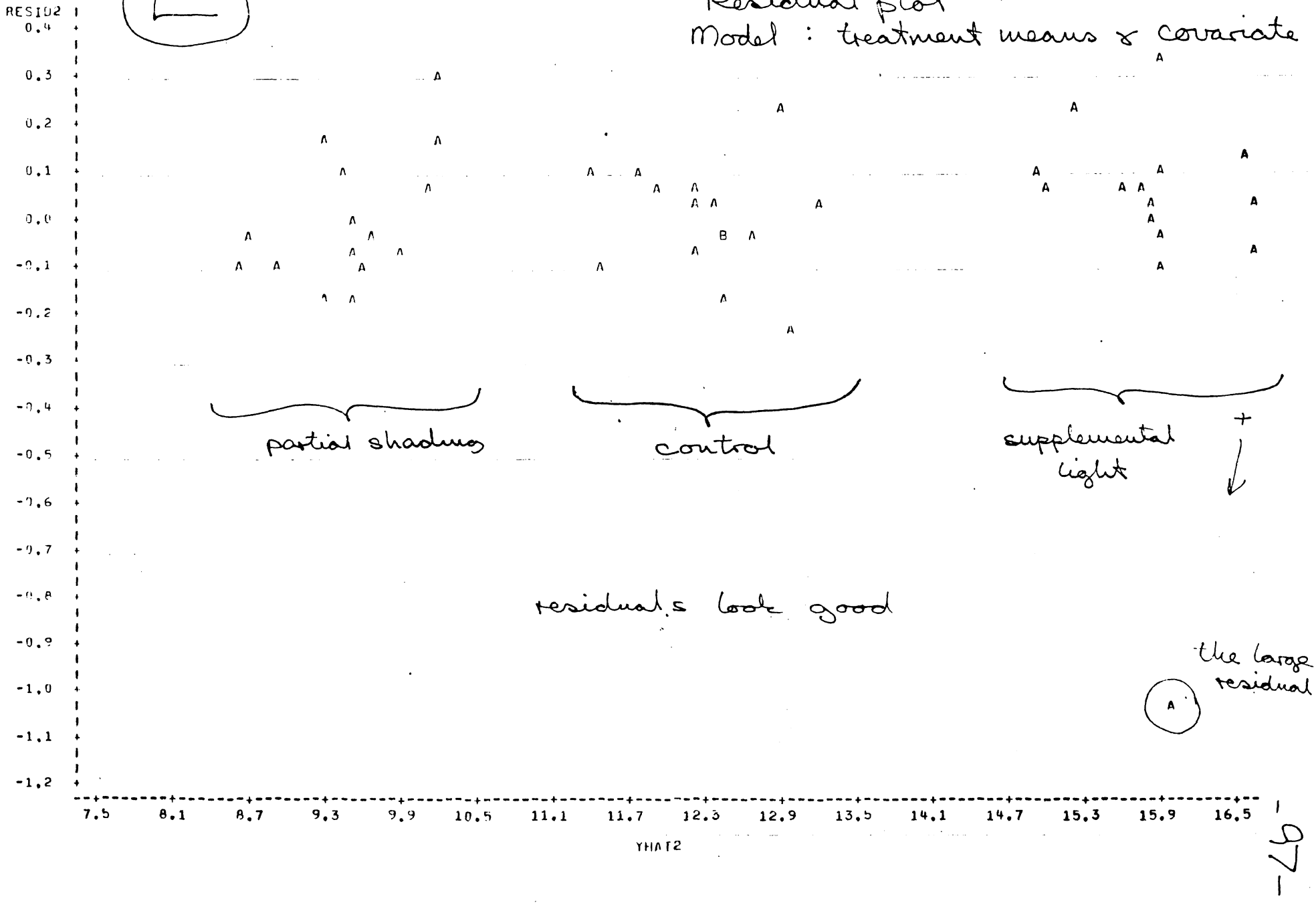


96

E

PLOT OF RESID2*YHAT2 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

Residual plot
Model : treatment means & covariate



residuals look good

A

the large residual

(F)

STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE Model : covariate & treatment means

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	3	313.09707581	104.33569194	2441.71	0.0001	0.994434	1.6463
ERROR	41	1.75195530	0.04273062		STD DEV		Y MEAN
CORRECTED TOTAL	44	314.75903111			0.20671385		12.55644444

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X4	1	R(4|0) =	1.80230171	42.18	0.0001	1	11.16913803	261.38	0.0001
X1	1		1.77236176	41.48	0.0001	0	0.00000000	.	.
X2	1	R(1,2,3|4) =	309.43241203	7241.47	0.0001	0	0.00000000	.	.
X3	0		0.00000000	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT for trt 3 =	6.32433507 B	30.99	0.0001	0.20410339
X4 b =	0.05776331	16.17	0.0001	0.00357318
X1 $\bar{y}_1(adj) - \bar{y}_3(adj)$ =	3.14572694 B	40.19	0.0001	0.07827463
X2 $\bar{y}_2(adj) - \bar{y}_3(adj)$ =	6.67728075 B	85.10	0.0001	0.07846687
X3	0.00000000 B	.	.	.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: 'BIASED ESTIMATOR' = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

→ or { intercept for trt 1 − intercept for trt 3
intercept for trt 2 − intercept for trt 3

10
8
1

STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

Model : 3 different regression lines

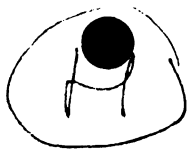
DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	313.12159433	62.62431887	1491.57	0.0001	0.994798	1.6319
ERROR	39	1.63743678	0.04198556		STD DEV		Y MEAN
CORRECTED TOTAL	44	314.75903111			0.20490378		12.55644444

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X1	1	different intercepts	1.25780444	29.96	0.0001	0	0.00000000	.	.
X2	1		300.58005333	7159.13	0.0001	0	0.00000000	.	.
X3	0		0.00000000	.	.	0	0.00000000	.	.
X4	1	covariate	11.16913803	266.02	0.0001	0	0.00000000	.	.
X5	1	different slopes	0.00369270	0.09	0.7684	0	0.00000000	.	.
X6	1		0.11082582	2.64	0.1123	0	0.00000000	.	.
X7	0		0.00000000	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT (for 3rd regr. line)	5.85980166	B	16.32	0.0001	0.35908468
X1 (intercept 1 - intercept 3)	3.68088132	B	7.77	0.0001	0.46795061
X2 (intercept 2 - intercept 3)	7.43092817	B	15.88	0.0001	0.46780016
X3	0.00000000	B	.	.	.
X4 (slope for 3rd tr =	0.06619465	B	10.28	0.0001	0.00644194
X5 tr1 slope - tr 3 slope =	-0.00985714	B	-1.12	0.2692	0.00879488
X6 tr2 slope - tr 3 slope	-0.01430931	B	-1.62	0.1123	0.00880778
X7	0.00000000	B	.	.	.

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: (BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.



STATISTICAL ANALYSIS SYSTEM

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	313.12159433	62.62431887	1491.57	0.0001	0.994798	1.6319
ERROR	39	1.63743678	0.04198556		STD DEV		Y MEAN
CORRECTED TOTAL	44	314.75903111			0.20490378		12.55644444

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
X4	1	slope ≠ 0 :	1.80230181 R(4|0)	42.93	0.0001	0	0.00000000	.	.
X1	1	means different	1.77236196 R(1|04)	42.21	0.0001	0	0.00000000	.	.
X2	1		309.46241203 R(2|014)	7369.97	0.0001	0	0.00000000	.	.
X3	0		0.00000000 R(3|024)	.	.	0	0.00000000	.	.
X5	1	slopes different	0.00369070 R(5|0-4)	0.09	0.7684	0	0.00000000	.	.
X6	1		0.11082582 R(6|0-5)	2.64	0.1123	0	0.00000000	.	.
X7	0		0.00000000 R(7|0-6)	.	.	0	0.00000000	.	.

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT	5.85980166 B	16.32	0.0001	0.35908468
X4	0.06619465 B	10.28	0.0001	0.00644194
X1	3.66089132 B	7.87	0.0001	0.46795061
X2	7.43092817 B	15.88	0.0001	0.46780016
X3	0.00000000 B	.	.	.
X5	-0.00985714 B	-1.12	0.2692	0.00879448
X6	-0.01430991 B	-1.62	0.1123	0.00880778
X7	0.00000000 B	.	.	.

Similar to the suggested analysis on p. 19-12, except that here there are 3 treatments and therefore 3 slopes.
ANOVA equivalent to table on bottom of p. 19-13

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS. THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT OF THE BIASED ESTIMATOR AND THE T VALUE TESTS H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE LETTER B ARE BLUE FOR THE PARAMETER.

OBSERVATION	OBSERVED VALUE	PREDICTED VALUE	RESIDUAL
1	11.88000000	11.79418327	0.08581673
2	12.76000000	12.97727092	-0.21727092
3	12.68000000	12.69558338	-0.01558338
4	11.96000000	11.90685828	0.05314172
5	12.40000000	12.35755834	0.04244166
6	12.24000000	12.41389585	-0.17389585
7	13.25000000	13.20262094	0.04737906
8	11.39000000	11.51249573	-0.12249573
9	12.45000000	12.47023335	-0.02023335
10	12.19000000	12.24486332	-0.05486332
11	13.18000000	12.92093481	0.25906519
12	11.46000000	11.39982074	0.06017926
13	12.37000000	12.41389585	-0.04389585
14	12.27000000	12.24488332	0.02511668
15	12.32000000	12.24488332	0.07511666
16	15.00000000	15.00292621	-0.00292621
17	16.68000000	16.55946838	0.12053162
18	16.59000000	16.55946838	0.03053162
19	15.09000000	15.10669569	-0.01669569
20	16.24000000	15.88496677	0.35503323
21	15.44000000	15.26234991	0.17765009
22	15.80000000	15.88496677	-0.08496677
23	15.88000000	15.83308203	0.04691797
24	16.00000000	15.88496677	0.11503323
25	15.68000000	15.62554308	0.05445692
26	15.82000000	15.83308203	-0.01308203
27	15.86000000	15.88496677	-0.02496677
28	14.98000000	15.98873625	-1.00873625
29	16.70000000	16.50758364	0.19241636
30	15.84000000	15.78119730	0.05880270
31	9.50000000	9.30192343	0.19807657
32	9.52000000	9.63289668	-0.11289668
33	9.51000000	9.50050738	0.00949262
34	9.45000000	9.50050738	-0.05050738
35	9.56000000	9.43431273	0.12568727
36	10.20000000	10.22864852	-0.02864852
37	8.82000000	8.83856089	-0.01856089
38	10.37000000	10.29484317	0.07515683
39	8.65000000	8.57378229	0.07621771
40	10.49000000	10.29484317	0.19515683
41	8.53000000	8.50756764	0.02241236
42	9.64000000	9.69909133	-0.05909133
43	9.38000000	9.56670203	-0.18670203
44	9.85000000	9.96386993	-0.11386993
45	9.17000000	9.30192343	-0.13192343

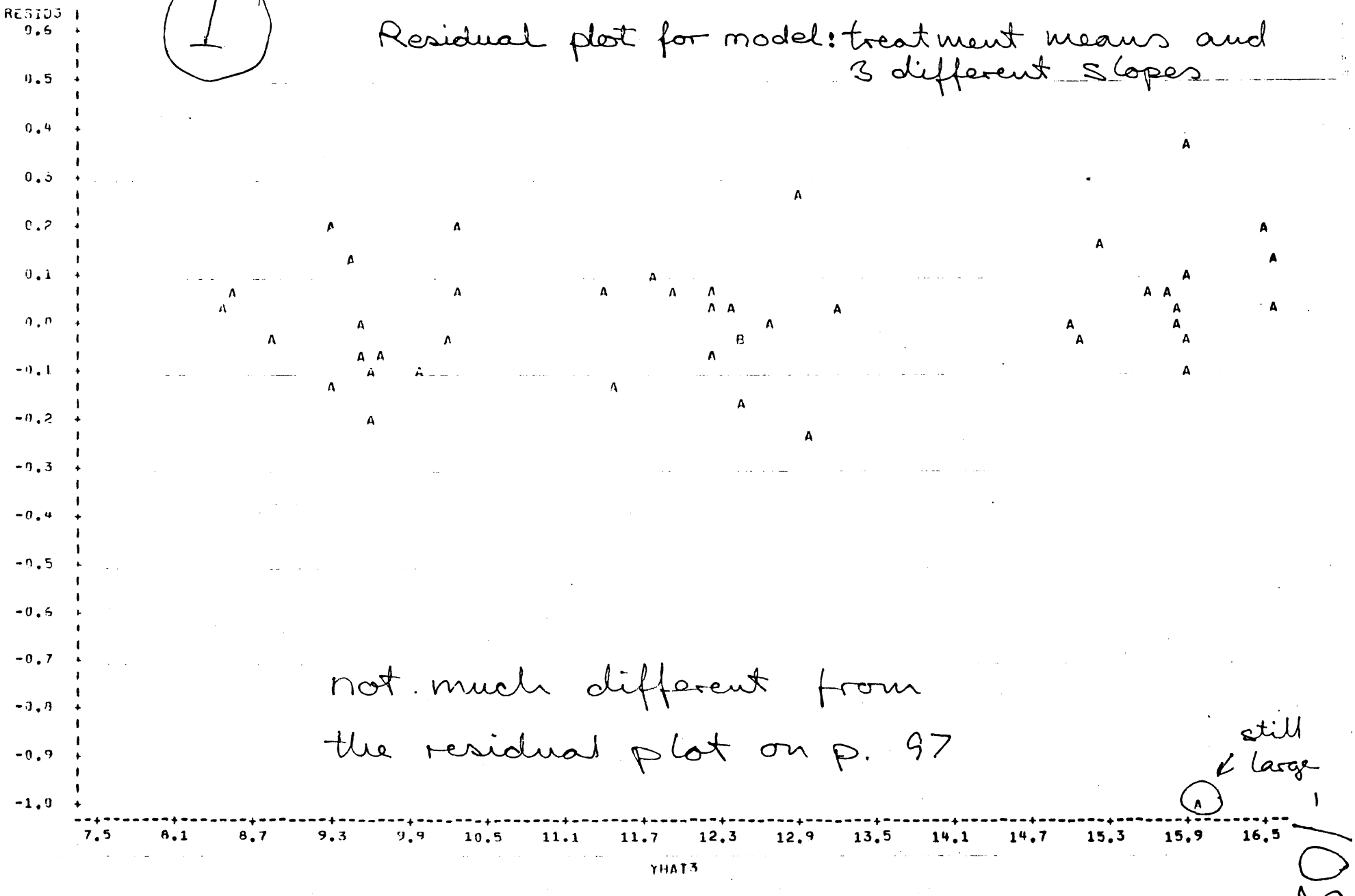
only large residual

SUM OF RESIDUALS	0.00000000
SUM OF SQUARED RESIDUALS	1.63743678
SUM OF SQUARED RESIDUALS - ERROR SS	0.00000000
FIRST ORDER AUTOCORRELATION	-0.05491787
DURBIN-WATSON D	2.09470947

PLOT OF RESID3*YHAT3 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

(I)

Residual plot for model: treatment means and 3 different slopes



not much different from the residual plot on p. 97

still large